



Research Report 1987

Identifying, Preparing and Evaluating Army Instructors

**Heidi Keller-Glaze
Jonathan Bryson
Ryan Riley
Jeffrey Horey
ICF International**

**William R. Bickley
Army Research Institute**

April 2016

**United States Army Research Institute
for the Behavioral and Social Sciences**

Approved for public release; distribution is unlimited.

**U.S. Army Research Institute
for the Behavioral and Social Sciences**

**Department of the Army
Deputy Chief of Staff, G1**

Authorized and approved:

**MICHELLE SAMS, Ph.D.
Director**

Research accomplished under contract
for the Department of the Army by

ICF International

Technical review by

Shala N. Blue, U.S. Army Research Institute
Christine DiFeliciantonio, Maneuver Center of Excellence Staff & Faculty Development
Branch

NOTICES

DISTRIBUTION: This Research Report has been submitted to the Defense Information Technical Center (DTIC). Address correspondence concerning ARI reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, Attn: DAPE-ARI-ZXM, 6000 6th Street Building 1464 / Mail Stop: 5610), Fort Belvoir, VA 22060-5610.

FINAL DISPOSITION: Destroy this Research Report when it is no longer needed. Do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this Research Report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE		<i>Form Approved OMB No. 0704-0188</i>		
1. REPORT DATE (DD-MM-YYYY) April 2016	2. REPORT TYPE Final	3. DATES COVERED (From - To) April 2012 – April 2013		
4. TITLE AND SUBTITLE Identifying, Preparing and Evaluating Army Instructors		5a. CONTRACT NUMBER W5J9CQ-11-D-0002		
		5b. GRANT NUMBER		
		5c. PROGRAM ELEMENT NUMBER 633007		
6. AUTHOR(S) Heidi Keller-Glaze, Jonathan Bryson, Ryan Riley, Jeffrey Horey; William R. Bickley		5d. PROJECT NUMBER A792		
		5e. TASK NUMBER 225		
		5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) ICF International 9300 Lee Highway Fairfax, VA 22031		8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U. S. Army Research Institute for the Behavioral & Social Sciences 6000 6 TH Street (Bldg. 1464 / Mail Stop 5610) Fort Belvoir, VA 22060-5610		10. SPONSOR/MONITOR'S ACRONYM(S) ARI		
		11. SPONSOR/MONITOR'S REPORT NUMBER(S) Research Report 1987		
12. DISTRIBUTION/AVAILABILITY STATEMENT: Distribution Statement A. Approved for public release; distribution is unlimited.				
13. SUPPLEMENTARY NOTES ARI Research POC: Dr. William R. Bickley, Fort Benning Research Unit				
14. ABSTRACT The Army Learning Model (ALM) calls for a re-examination of instructor selection and training. Because the ALM is learner-centric, it specifies that instructors must now become facilitators in a more distributed classroom role. As facilitators, they must acquire skills at tailoring instruction to learners' personal characteristics and at employing technology-enabled learning tools. Although the ALM outlines the end-state of the re-examination of instructors, it does not directly address the processes by which the Army is to attain the end-state. This report addresses the need to better explicate the processes by which the Army can select, train, and assess instructors in support of the ALM. General instructor selection, preparation, and assessment processes are addressed. From this, an operational definition of an effective Army instructor is derived, as are the KSAOs for instructors. A framework for the Army's utilization of the KSAOs in instructor selection, preparation, and assessment is provided.				
15. SUBJECT TERMS Instructors, Training, Selection, Assessment				
16. SECURITY CLASSIFICATION OF:		17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Dr. Scott E. Graham
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified	Unlimited Unclassified	78
19b. TELEPHONE NUMBER 706-545-2362				

Research Report 1987

Identifying, Preparing and Evaluating Army Instructors

**Heidi Keller-Glaze
Jonathan Bryson
Ryan Riley
Jeffrey Horey
ICF International**

William Bickley
Army Research Institute

**Fort Benning Research Unit
Scott E. Graham, Chief**

April 2016

Approved for public release; distribution is unlimited.

IDENTIFYING, PREPARING AND EVALUATING ARMY INSTRUCTORS

EXECUTIVE SUMMARY

Research Requirement

The Army Learning Model (ALM) (US Army Training & Doctrine Command, 2011) calls for a re-examination of the Army instructor role. Being learner-centric, ALM specifies that instructors must now become facilitators in a more distributed classroom role. As facilitators, they must acquire skills at tailoring instruction to learners' personal characteristics and at employing technology-enabled learning tools. Although the ALM outlines the end-state of the re-examination of instructors, it does not directly address the processes by which the Army is to attain that end-state. Army Learning Model requirements. To address this gap, this effort examined current practices in instructor selection, preparation, and assessment as they might be applied to instructor transition under ALM.

Procedures

Based on a literature review of effective instructors, including teachers, trainers and facilitators, a draft set of job and person requirements for an effective instructor was compiled. From the initial knowledge areas, skills, abilities, other characteristics (KSAOs) and work behaviors associated with effective instructors, an operational definition of an effective instructor for Army training was developed and a list of recommended KSAOs and work behaviors that describe effective instructors was compiled.

Working from the definition, KSAOs, and work behaviors, reviews of the instructor/trainer/teacher literature related to identification and selection, training and preparation, and assessment and evaluation methods were conducted, focusing on post-secondary literature from the previous 10 years. Pertinent findings for selection, preparation and evaluation methods were documented as were recommendations for each method.

Also, for each KSAO and work behavior, we analyzed during which process, selection or preparation, the Army could assess instructors for that KSAO and work behavior

Results

The initial review resulted in an operational definition of an effective Army instructor and eight knowledge areas, nine skills, six abilities, and nine other characteristics determined to be necessary for instructor to achieve 13 instructor effectiveness work behaviors. The literature review on best practices in teacher and instructor selection, training and development and evaluation provided specific methods and the empirical support for these methods in use to better ensure instructor effectiveness. From these tasks, a framework was constructed that describes which process, selection, preparation or evaluation, was best suited for ensuring instructors possess the critical KSAOs and work behaviors. In addition, specific methods and techniques within each process were identified to better inform Army instructor selection, preparation and evaluation across a broad range of learning contexts.

Utilization and Dissemination of Findings

This report provides information suitable for reconsidering evaluation dimensions and measurement for selection, preparing, and assessing Army instructors, facilitators, and coaches. The information may be considered supplemental to current Army instructor effectiveness doctrine, such as TR 600-21 which focuses on non-commissioned officer selection, training and education, and assessment policies and procedures. The current KSAOs, work behaviors and recommended methods for selecting, preparing and evaluating instructors are focused on best practices from the empirical literature and incorporate ALM concepts and constructs, particularly enhancing learner-centered aspects of institutional training.

IDENTIFYING, PREPARING AND EVALUATING ARMY INSTRUCTORS

CONTENTS

	Page
INTRODUCTION.....	1
METHODS.....	2
Foundational Task.....	2
Instructor Selection.....	3
Instructor Preparation.....	4
Instructor Evaluation.....	4
Considerations for Instructor Selection, Preparation, and Evaluation.....	5
RESULTS	
Foundational Task	
Instructor Selection Methods.....	8
Instructor Preparation Methods.....	16
Instructor Evaluation Methods.....	24
Considerations for Instructor Selection, Preparation, and Evaluation.....	40
DISCUSSION.....	48
Operational Definition of an Effective Instructor.....	48
Effective Methods for Selecting Instructors.....	49
Effective Methods for Preparing Instructors.....	50
Effective Methods for Evaluating Instructors.....	50
Considerations for Instructor Selection, Preparation, and Evaluation.....	53
Overall.....	53
REFERENCES.....	54

LIST OF TABLES

	Page
TABLE 1. WORK BEHAVIORS OF AN EFFECTIVE INSTRUCTOR.....	6
TABLE 2. KNOWLEDGE REQUIRED FOR AN INSTRUCTOR TO BE EFFECTIVE.....	7
TABLE 3. SKILLS REQUIRED FOR AN INSTRUCTOR TO BE EFFECTIVE.....	7
TABLE 4. ABILITIES REQUIRED FOR AN INSTRUCTOR TO BE EFFECTIVE.....	8
TABLE 5. OTHER CHARACTERISTICS REQUIRED FOR AN INSTRUCTOR TO BE EFFECTIVE	8

TABLE 6. TR 600-21 INSTRUCTOR ASSESSMENT MATERIALS.....	28
TABLE 7. KSAOs AND WORK BEHAVIORS LINKED TO APPROPRIATE PROCESSES.....	41
TABLE 8. EXAMPLES OF INSTRUCTIONAL STRATEGIES AND TECHNIQUES AND POSITIVE STUDENT OUTCOMES.....	50
TABLE 9. EMPIRICAL SUPPORT FOR IDENTIFICATION/SELECTION METHOD.....	51
TABLE 10. ASSESSMENT/EVALUATION METHODS' APPLICABILITY TO FACETS OF INSTRUCTOR EFFECTIVENESS.....	54

ARMY INSTRUCTOR IDENTIFICATION, PREPARATION AND EVALUATION

INTRODUCTION

The Army is currently facing, and will face for the foreseeable future, several challenges in terms of the way it trains and educates Soldiers. Among these challenges and their implicit requirements are:

- High operational tempo that requires maximizing the efficiency of time spent training and educating Soldiers.
- Varied and rapidly changing operational environments, requiring lessons learned in the field to be incorporated quickly into training and education.
- Technologically savvy Soldiers entering the Army, many of whom are very comfortable with or even dependent on technology for learning and staying connected to others.
- Soldiers with great depths of real world experience gained from repeated deployments to the conflicts in Iraq and Afghanistan who wish to share this experience.
- Evolving requirements for instruction and expansion of the construct of “instructors” to meet the challenges identified.

The implications of these challenges are many: time spent training and educating Soldiers must be used efficiently; opportunities to train and educate must be maximized (i.e., not all training can take place in a classroom); content and materials used in training and education must be adaptable and incorporate lessons learned in as near to real-time as possible; and training and education must not only be engaging for experienced Soldiers, but must also make use of their wealth of experience.

As part of the response to these challenges, the Army is exploring how to effectively shift from instructor-centric models of training and education to learner-centric models of training and education. These are central themes of the United States Army Learning Model (ALM, U.S. Army Training and Doctrine Command, 2011). The ALM directs course proponents to 1) use more problem-solving approaches in classrooms, where the instructor takes on more of a facilitator role rather than lecturer role, and 2) make training and education more learner-centric by customizing content and methods/modalities to the learner’s needs and leverage the learners’ wealth of experience. These imply that the instructor role in Army education is shifting from a traditional lecture approach to one that supports more student-centric, problem-based training, education and professional development.

Although the ALM outlines the end-state of the role of instructors, it does not directly address the processes by which the Army is to attain the end-state. More specifically, the ALM leaves open the processes by which the Army should select, train, and assess instructors in support of the ALM. This report investigates these three inter-related processes and, from its findings, outlines considerations for the transition to facilitative, learner-centric instructors.

To accomplish this objective, the effort included five tasks:

1. Specify the instructor job and person requirements and develop an operational definition of an effective instructor.
2. Review the current practices and considerations in instructor selection.
3. Review the current practices and considerations in instructor preparation.
4. Review the current practices and considerations in instructor assessment.
5. Develop considerations for the Army to use in selecting, preparing, and assessing instructors.

The remainder of this document is divided into “Methods” and “Results” sections and a summary “Discussion” section. The Methods and Results sections are both divided into five subsections with the subsections corresponding to the five tasks above. The five Methods subsections give, for each of the tasks, an overview of the procedures used to complete the task. The five Results subsections then give, for each of the tasks, the findings for that task. Finally, the discussion section contains overall summary conclusions drawn across the five tasks.

METHODS

Foundational Task

To provide a foundation for the project, an abbreviated job analysis was conducted to identify and document instructor job and person requirements. This analysis was performed in three steps.

- Initial literature review
- Workshop
- Expanded literature review

Initial literature review. An initial review of the military and civilian education and training literature was conducted to identify job and person requirements for instructors to be effective.

Seven primary research questions were used to help guide this effort:

- What work behaviors do instructors need to perform to be effective in the classroom?
- What knowledge do instructors need in order to be effective in the classroom?
- What skills do instructors need to be effective in the classroom?
- What abilities do instructors need to be effective in the classroom?
- What other characteristics do instructors need to be effective in the classroom?
- What does it mean to be an effective instructor?

Relevant military and civilian data sources were identified and reviewed. This initial literature review identified 25 articles that yielded a draft set of job and person requirements for an effective instructor.

Workshop. To refine the draft set of job and person requirements identified during the literature review, a three hour workshop was held with subject matter experts (SMEs). The workshop was conducted in four segments. First was a broad topic discussion of what it means to be an effective instructor. Then participants were shown the list of job requirements identified during the literature review and asked a series of open-ended questions regarding the list's accuracy and what might change with differences in type of training. Next, a similar set of open-ended questions was asked of participants regarding the list of person requirements developed during the literature review. Last, a brainstorming session was conducted to identify additional literature sources as well as experts within the field of education and training that could help with later phases of the project.

Expanded literature review. Based on the information and guidance provided by the workshop, the lists of job and person requirements for an effective instructor were revised and improved. These lists were further refined through a more extensive review of the education and training literature and through follow-up discussions with several subject matter experts (SMEs) identified in the workshop.

Definition of an effective instructor. Reiteratively drawing on the workshop discussions and the refined lists of job and person requirements (along with additional information found in the education and training literature), an operational definition for what it means to be an effective instructor was posited.

Instructor Selection

A systematic review of the military, academic and industry literature was conducted to explore and understand the initial qualifications and relevant criteria used in identifying and selecting instructors,

Five primary research questions guided the identification effort. These included:

- What qualifications and methods are currently being used in the Army to select instructors?
- What qualifications and methods are currently being used in academia to select instructors?
- What qualifications and methods are currently being used in industry to select instructors?
- What empirical support do current selection practices have in the research literature?
- What does the literature suggest are the most effective methods for selecting instructors?

Content from a range of data sources was reviewed and evaluated during the literature review. Also, external subject matter experts in the field of education and training were solicited for articles and/or references on instructor identification and selection.

Instructor Preparation

We conducted a literature review of the empirical and conceptual literature related to instructor preparation, development and certification processes. This resulted in organizing this literature into four sections:

- research on instructional method effectiveness
- recent military instructional and certification research
- army instructor development and certification practices
- other military instructor development and certification practices.

Included is information related to the preparation of students other than instructors as this is where the bulk of the empirical findings of the effects of learning methods on student and other outcomes are focused. This literature was deemed to be pertinent to the effort as instructors are a subset of all students and training participants and, indeed, effective instructors should be aware of effective instructional methods.

This review focused on methods of teaching, training and instructing as opposed to methods and principles of learning. Methods of learning include reading, observing and other forms of self-study. However, it is impossible to completely distill methods of instruction from methods of learning, as methods of learning play a role in instruction effectiveness. In each section, an attempt is made to tie the relevance of the research to Army instructor effectiveness.

Instructor Evaluation

This review identified effective methods for instructor assessment and evaluation across settings to inform guidance for use in the Army, including both formal and informal learning environments. A review of academic, military and empirical support for instructor assessment and evaluation methods was conducted. Five primary research questions guided this effort::

- What methods are currently being used in the Army and other Services to assess and evaluate instructors?
- What methods are currently being used in academia to assess and evaluate instructors?
- What methods are currently being used in industry to assess and evaluate instructors?
- What empirical support do current assessment and evaluation practices have in the research literature?
- What does the literature suggest are the most effective methods for assessing and evaluating instructor effectiveness?

Considerations for Instructor Identification, Preparation and Evaluation

A framework for optimal methods of Army instructor identification, preparation and evaluation was built upon the literature review findings integrated with project team recommendations.

Following the results of the foundational task that identified pertinent Army instructor KSAOs and work behaviors, the project team participated in a judgment exercise to assign the phase (selection, preparation, or assessment) at which to ensure instructors possess each of the KSAOs. Work behaviors were associated with the KSAOs to add additional context for making the judgments. Five project team members familiar with the foundational task, the literature review findings and Army instructor job requirements provided judgments. The team then participated in a group working session in which discrepancies in assignment were discussed and a consensus agreement on the proper phase was determined for each KSAO.

In addition to making a judgment of the phase at which to determine instructors possess the KSAO, different implementation methods were also proposed for the phase selected. This information was considered valuable regarding the potential selection or development of instruments, methods and techniques applicable to each KSAO. For example, for the instructor KSAO of “Skill at observing and monitoring students,” the preparation phase was selected to be optimal, with Reading, Lecture, Problem Solving, Role Play, and Simulation the methods identified as potentially valuable for skill development.

RESULTS

Foundational Task

The results of the abbreviated job analysis are organized below into three main sections. The first section displays the job requirements (i.e., work behaviors) instructors need to exhibit in order to be effective. The second section provides the person requirements (i.e., KSAOs) instructors need to have in order to be effective. The final section provides the operational definition for an effective instructor that was developed from the results of the job analysis.

Work behaviors. A total of 13 job requirements (i.e., work behaviors) were identified as essential for instructors to exhibit in order to be effective (see Table 1).

Table 1

Work Behaviors of an Effective Instructor

No.	Description
WB1	Monitor/observe students to ensure learning is taking place and that problems/issues (e.g., learning off track, faulty thinking) are identified and addressed.
WB2	Evaluate student performance to determine if they are progressing and meeting the general outcomes and specific objectives of the course.
WB3	Use frequent practical exercises, exams, and other assessment techniques.
WB4	Plan/prepare lessons and activities to achieve learning objectives, maximize student potential, and address different learning preferences of students.
WB5	Maintain expertise in topic areas to better facilitate/guide student learning in the classroom.
WB6	Build rapport with students to ensure they are engaged in learning, feel comfortable asking questions, have a positive affect towards both the instructor and the material being taught and to develop credibility with students.
WB7	Select/implement instructional strategies and techniques to account for differences in subject domain/content, the learning environment, and individual differences in student behaviors and thought processes.
WB8	Communicate information and ideas orally so others will understand.
WB9	Communicate information and ideas in writing so others will understand.
WB10	Accurately and effectively interpret students' comments in both verbal and written form.
WB11	Apply learning theory to individual instructional circumstances.
WB12	Combine pieces of information to form general rules and conclusions (includes finding a relationship among seemingly unrelated events).
WB13	Apply general rules to specific problems to produce answers that are reasonable.

KSAOs A total of 32 KSAOs were identified as essential for instructors to be effective (see Tables 2, 3, 4 and 5). Knowledge elements spanned specific course content (K1), general teaching and evaluation strategies and methods (K4, K5, K6, K8), learner characteristics (K2, K3), and communication (K7). Similarly, the skills also focused on application of knowledge in student observation and assessment (S1, S2, S4), teaching and coaching strategies (S3, S5, S7, S8), providing feedback (S6) and using technology (S9). Abilities focused on communication and information organization (A1, A2, A5), interpreting student inputs (A3), and applying learning to specific circumstances and problems (A4, A6). Lastly, other characteristics spanned a variety of traits, perspectives and values that would contribute to being open to student needs and professional development and behaviors.

Table 2

Knowledge Required for an Instructor to be Effective

No.	Description
K1	Subject matter being taught, in order to utilize content knowledge effectively, determine student mastery of content, and better facilitate/guide student learning.
K2	Traits and behaviors of adult learners, in order to perceive when students have or have not gained mastery of the material and/or are capable of transitioning to new materials and/or methods.
K3	Students' current level of performance, in order to evaluate when material/tasks are precisely challenging enough.
K4	Principles and methods for curriculum and training design in order to align course goals and learning objectives with intended student outcomes.
K5	Principles and methods of teaching individuals and groups in order to accommodate and address the different learning needs of each student.
K6	Principles and methods for assessing for training effectiveness in order to ensure students are progressing and that learning is taking place.
K7	The structure and content of the English language, in order to effectively facilitate and present information.
K8	Coaching methods and techniques to facilitate student motivation and social learning.

Table 3

Skills required for an Instructor to be Effective

No.	Description
S1	Observe and monitor students in order to assess whether knowledge has been transferred or content has been mastered.
S2	Employ questioning techniques (e.g., probing, open-ended questioning) to facilitate discussion and/or assess student knowledge transfer and content mastery.
S3	Utilize techniques such as summarizing and reiterating to ensure accurate interpretation, clarify student level of learning, and elaborate upon student ideas.
S4	Formal and informal assessment to evaluate student progress on core course content.
S5	Make use of multiple instructional strategies and techniques (e.g., scaffolding, blended learning) to account for differences in subject domain/content, the learning environment, and individual differences in student behavior/thought processes.
S6	Provide formal and informal feedback so students can recognize strengths and weaknesses and determine how to improve performance.
S7	Present and facilitate course materials so that content and learning objectives are sequenced appropriately and to bring students to an end goal reflective of the original course design.
S8	Mentor and coach students to help them achieve course objectives, diagnose weaknesses, and continually improve performance.
S9	Apply educational technology in ways that enhance student learning.

Table 4

Abilities required for an Instructor to be Effective

No.	Description
A1	Communicate information and ideas orally so others will understand.
A2	Communicate information and ideas in writing so others will understand.
A3	Accurately and effectively interpret students' comments in both verbal and written form.
A4	Apply learning theory to individual instructional circumstances.
A5	Combine pieces of information to form general rules and conclusions (includes finding a relationship among seemingly unrelated events).
A6	Apply general rules to specific problems to produce answers that are reasonable.

Table 5

Other Characteristics required for an Instructor to be Effective

No.	Description
O1	Have openness to experience or a high degree of intellectual curiosity, creativity, and preference for novelty and variety.
O2	Have low need for control and tolerance for ambiguity to allow for classroom discussion and group problem-solving when applicable.
O3	Believe students are responsible for and capable of own learning.
O4	Value independent thought.
O5	View learning as a collaborative process to enable student participation.
O6	Accept student-centered methods (e.g., experiential learning, case-based learning, inquiry-based learning) as valid.
O7	View teaching as a learning profession.
O8	Is content with one's current life situation.
O9	Is highly persistent and passionate toward achieving long-term goals.

Instructor Selection Methods

Instructor selection methods in the Army. There are several Army regulations that provide guidance on how to identify and select non-commissioned officers for instructor duty. These include: Army Regulation 614-200, Enlisted Assignments and Utilization Management, U.S. Army Training and Doctrine Command (TRADOC) Regulation 600-21, and TRADOC Regulation 350-70, Army Learning Policy and Systems.

Army Regulation 614-200. Army Regulation (AR) 614-200 is the regulation governing the “selection of enlisted Soldiers for assignment, utilization, reclassification, detail, transfer, and

training..." (U.S. Department of the Army, 2011, p.i). Section 2 of Chapter 6 lays out the basic qualifications required for selecting non-commissioned officers for instructor duty. These qualifications span a range of criteria to include having a high school diploma or GED equivalent to passing the APFT to not having a speech impediment.

In addition to these basic criteria, there are further prerequisites for instructors, depending on where they will be assigned (e.g., Sergeants Major Academy, Uniformed Services Schools, Basic Officer Leaders Course (BOLC), or Army Reserve Officers Training Corps). Such prerequisites include being a graduate of the course they will be teaching (e.g., be a Senior Leaders NCOES course graduate if SFC or MSG) and prior experience in certain duty positions (e.g., served in principal duties of primary military occupational specialty within last 2 years).

TRADOC Regulation 600-21. TRADOC Regulation 600-21 governs the implementation of the non-commissioned officer education system (NCOES) instructor development and recognition program (U.S. Army Training and Doctrine Command, 2013). It indicates that, where possible, evidence-based selection processes should be used to select instructors for the non-commissioned officer academies and suggests a two phase procedure.

For phase I, Soldiers submit an instructor application packet to the Non-commissioned Officer Academy (NCOA) (see Appendix B of TRADOC 600-21). The NCOA screens the packet for initial eligibility. Those Soldiers who are selected participate in a structured interview with representatives from the NCOA (i.e., commandant and at least one other person). The recommended protocol for these interviews is the Teacher Quality Index – Military (TQI-M) which evaluates a Soldier on 12 indicators of instructor effectiveness taken from Stronge & Hindman's (2006) original TQI. These indicators are organized into five categories of quality to include the instructor as a person, classroom management and organization, planning for instruction, implementing instruction, and monitoring student progress and potential.

Instructor selection methods in academia. Most selection methods within post-secondary education and public schools have been designed under the assumption that applicants have already self-identified as instructors and, in most cases, have completed coursework in the field of education

Post-secondary instructor selection. The Bureau of Labor Statistics' Occupational Outlook Handbook suggests the qualifications used in selecting post-secondary instructors for academic positions primarily consist of some combination of education, experience and certification. These requirements can vary widely depending on the type of institution (e.g., university, colleges, and trade schools). However, in general these often constitute the minimum requirements for most instructor positions. A recent examination of job listings in this area appears to support this assertion (The Chronicles of Higher Education, n.d.). For instance, a recent job announcement for an assistant professor of optometry required candidates to have a doctoral degree in Optometry, residency certification in ocular disease, be licensed to practice in the state where the position was located with Therapeutic Pharmaceutical Agent (TPA) certification, and be able to demonstrate clinical teaching abilities. Another job announcement for a welding instructor required candidates to have either a bachelor's degree with two years of experience as a welder or an associate's degree and six years of experience as a welder. They were also required to have a community college credential from the state where the position was

located authorizing service as a welding instructor.

In addition to basic education, experience, and certification requirements, many institutions also require candidates to demonstrate additional knowledge, skills, and abilities (e.g., ability to work well with others) as well as provide different types of documentation. For example, a recent job announcement for an assistant professor of construction management required candidates to have "...strong communication skills in spoken and written English." Another job announcement for an industrial and commercial electrical faculty required candidates to demonstrate their ability in working with culturally diverse populations. Both of these announcements had various documentation requirements as well, such as providing copies of transcripts, resumes/curricula vitae, and letters of reference.

The hiring methods used for selecting candidates for post-secondary instructor positions in academia tend to follow a similar pattern. In most cases, the selection process starts with the submission of an application packet that includes a range of different documentation (e.g., resumes, references). A hiring/search committee then evaluates and screens the application packet to ensure the candidate meets the minimum and preferred requirements of the position. Those candidates who are selected are then interviewed. In some instances, multiple interviews may be conducted. For example, some institutions may first conduct a phone interview with a candidate before scheduling an in-person interview. In addition to being interviewed, candidates may be asked to provide teaching demonstrations/presentations or provide impromptu writing samples. At the conclusion of the interview process, the hiring/search committee selects the candidate for the position.

Primary and secondary education instructors. The Bureau of Labor Statistics identified three basic prerequisites every state requires of its instructors in order to teach at the primary (i.e., kindergarten and elementary) and secondary (i.e., middle and high) levels of education. These include: 1) having at least a bachelor's degree, 2) completing a teacher preparation course, and 3) becoming certified to teach. The following section describes the requirements for each level of education in more detail.

At the kindergarten and elementary school levels, all states require instructors to have at least a bachelor's degree in elementary education and in some cases major in a particular content area such as math or science. Instructors at the middle school level have similar education requirements. However, some states require instructors to major in elementary education; while others require instructors to major in a particular content area. At the high school level, most states require instructors to have a bachelor's degree in the subject they will teach.

All states require instructors at both the primary and secondary levels of education to complete a teacher preparation program and supervisory teaching experience.

All states require instructors to be certified in the grade levels they plan to teach. For example, kindergarten and elementary school teachers are certified to teach early childhood grades; whereas, high school teachers are certified for secondary or high school grades. Certification requirements vary by state. However, most require instructors to pass both a general teaching certification test as well as a knowledge test on the specific content area being

taught. In addition, instructors at all levels are required to pass a background check before being employed.

In addition to these three prerequisites, many states also use additional knowledge, skills, and abilities. For example, a recent job announcement for a kindergarten teacher required candidates to be good at solving problems, communicating in writing, being professional, and working well with others (Teach.org, n.d.). Another job announcement for a middle school English language arts teacher listed several skills and traits that a successful candidate for the position should exhibit such as, "...a record of producing dramatic student achievement gains..., and a commitment to creating a structured, predictable and joyful environment for children." Most of these types of positions require documentation such as portfolios and resumes.

In terms of hiring practices, most public school systems appear to use one of two methods when selecting primary and secondary instructors. The first is similar to the process used to select post-secondary instructors. Specifically, candidates submit an online application (to include supporting documentation) to the hiring entity (e.g., school, district, system). The application is then screened to ensure the candidate meets the minimum qualifications for the job. Those who pass screening are called in for interviews. In some cases, candidates might undergo multiple interviews. For example, some public school systems use phone interviews as an additional screening mechanism before bringing candidates in for in-person interviews. Once a candidate has been interviewed, the hiring entity then makes a decision as to whether they should or should not hire the individual for the position.

The second method of selecting primary and secondary instructors is similar to the first. However, candidates do not submit an application for a specific job announcement, but they submit a general application that is then screened to ensure candidates meet the minimum requirements for being a teacher. Candidates who pass this screening are placed in a database of qualified applicants. This database is then used by schools to select candidates whose qualifications align with the needs of a job vacancy in their organization. Selected candidates are then called in to be interviewed. Once a candidate has been interviewed, the school makes a decision as to whether the individual should be hired.

Instructor selection methods in industry. A review of job announcements posted on the American Society of Training and Development (ASTD) website and LinkedIn suggest the primary criteria used when selecting instructors for work in corporate setting is education and experience level. For instance, a recent job announcement for a training specialist required successful candidates to have a minimum of five years progressive experience in learning and development, two to four years' experience training in a financial services call center, and a bachelor's degree in a related field. Another job announcement for a sales trainer called for a bachelor's degree, a minimum of two years training experience, and experience developing and facilitating adult training programs.

In addition to education and experience, many companies require instructors to demonstrate different knowledge, skills, and abilities, such as demonstrating knowledge of company technical processes or exhibiting a professional work ethic. They also tend to require different certifications depending on the area of focus for the instructor. For example, a recent

job announcement for a training specialist at an insurance firm required candidates to be certified in general insurance. Many also mentioned preference for candidates with a Professional in Human Resources (PHR), Senior Professional in Human Resources (SPHR), Certified Professional in Learning and Performance (CPLP) or similar certification.

The hiring processes used in selecting candidates for instructor positions in industry tend to follow a similar pattern. In most cases, the process starts with submission of an application packet that includes a range of different documentation depending on the position (e.g., resumes, references). The company then screens the application packet to ensure the candidate meets the requirements of the position. Those candidates who are selected are then interviewed. In some instances, multiple interviews may be conducted. For example, some institutions may first conduct a phone interview with a candidate before scheduling an in-person interview. At the conclusion of the interview process, the company selects the candidate(s) for the position.

Empirical support for current selection practices. Very little objective research is available on how effective instructors should be selected (Mertz, 2010). According to Guarino, Santibanez, & Daley (2006) this is a result of two primary factors. First, there is no agreed-upon definition of teacher quality among researchers. Second, the availability of data sources researchers can use to explore and identify the effective selection of teachers is limited. In fact, a review of the literature would suggest no complete set of empirically-tested skills, attitudes, interests, or abilities that consistently predict a teacher's effectiveness or the degree to which they successfully produce a desired outcome (e.g., student gains) in the classroom (Wise, Darling-Hammond, & Berry, 1987).

While a complete set does not currently exist, several characteristics and qualifications have been explored individually in the literature.

Intelligence. Research on teacher intelligence and its relation to teacher effectiveness is not a new area of interest. In fact, studies go back as far as the early 1900s (Darling-Hammond, 1999). Since that time a variety of different proxy measures of teacher intelligence have been explored (McEachin & Brewer, 2011). These have included: general measures of intelligence (Morsh & Wilder, 1954), standardized test scores (D'Augostino & Powers, 2009; Gimbert & Chelsey, 2009; Schalock, 1979; Webster, 1988; Wise et al., 1987) and academic ability (Heinz, 2013; Schalock, 1979; Wise et al., 1987;). The following section discusses the literature on each of these in further detail.

In 1954, the Air Force Personnel and Training Research Center at Lackland Air Force Base examined 55 correlational studies that assessed the relationship between teacher intelligence and teacher effectiveness (Morsh & Wilder). These studies assessed teacher intelligence using a variety of different intelligence examinations (e.g., American Council on Education Psychological Exam, Army Alpha exam). Teacher effectiveness was assessed either by ratings or rankings of teacher performance (e.g., student, peer, administrative), observations of teacher behavior in teaching situations, or student gains. Results from this effort were mixed with student gains having the highest correlation with teacher intelligence. (Morsh & Wilder, 1954).

Twenty five years later the idea of a comprehensive evaluation of teacher intelligence literature was revisited by Schalock (1979). In addition to reviewing Morsh and Wilder's (1954) findings, he looked at several new studies that used other proxy measures of teacher intelligence to include: college entrance exam scores, teacher certification scores, and college grade point averages. Results from these efforts were again mixed. For example, in one study, no correlation was found between college entrance exam scores, certification scores (i.e., National Teacher Examination), and teacher effectiveness (i.e., principals' ratings and pupil achievement). They did, however, find a small positive correlation between a teacher's grade point average (GPA) and both criteria. In another study, they found a negative (albeit not statistically significant) relationship between teachers' GPAs and their performance (i.e., principal ratings) during their first year of teaching. There was also no relationship between a teacher's Scholastic Aptitude Test (SAT) scores and their performance as a teacher.

In 2003, a systematic review of studies on the relationship between teacher characteristics and student performance (e.g., student standardized test scores, graduation rates, attendance at postsecondary institutions, and acquisition of knowledge and skills not easily measured by standardized tests) was conducted by Wayne and Youngs. As part of that effort, they reviewed seven studies involving proxy measures of teacher intelligence. These included: licensure exam scores (2 studies), verbal skills scores (3 studies), college entrance exam scores (1 study), and a single multiple choice mathematics item score (1 study). Results were mixed. Of the two studies that used licensure exams as their predictor, one found that higher scores on the NTE Common examination had a negative or indeterminate relationship with student performance depending on their education level (i.e., elementary, secondary). The other used the Texas Examination of Current Administrators and Teachers (TECAT) and found a positive relationship between teacher scores on the exam and student test scores in reading. As for the three studies that looked at teacher scores on a verbal skills test, they found that either no relationship existed between the two constructs or if there was one, it was small and positive. The study that used the single multiple choice mathematics test item found that teachers who answered the question correctly had larger math gains in classroom than those that answered it incorrectly. A study that used composite scores (i.e., English, mathematics, social studies reading, natural science reading) from the American college testing (ACT) entrance exam, found a positive relationship between a teacher's scores and reading gains by elementary school student; however, this was not the case for student math scores at that level.

More recently, in 2011 McEachin and Brewer conducted a review of the teacher intelligence literature that focused on several current achievement and licensure tests being used today. They found that much of the literature painted a confusing picture. For example, the Praxis was found to have only a small positive relationship with student math scores and no relationship with student English scores. The California licensure exams (i.e., California Subject Examinations for Teachers, California Basic Education Skills Test, and the Reading Instruction Competence Assessment) were found to have no significant relationship with increases in student English language acquisition or math achievement.

Level of education. A person's education level and education's impact on job performance (not limited to teachers) is another topic that has spurred interest for many years. A comprehensive meta-analysis of 293 empirical studies on this relationship (Ng & Feldman,

2009) found a positive relationship between education level and performance of job duties and organizational citizenship behavior performance. They also found a negative relationship between education level and counterproductive behavior performance.

Given these results, it would appear level of education should be an important predictor of overall instructor performance. However, much education literature seems to tell a different story. According to Kane, Rockoff, and Staiger (2006), “the literature on teacher effectiveness has consistently failed to find that those holding master’s degrees are more effective, despite the fact that most teacher pay scales reward higher educational attainment” (p.22). In 2003, Wayne and Youngs reviewed the available research to date in this area and found that most studies were inconclusive on the relationship. Those that did show a connection between education level and performance were unreliable, with some studies showing a positive relationship and others showing a negative relationship. Wayne and Youngs (2003) also reviewed studies that looked into the specific degree held by the teacher (e.g., math, science). Results again were generally inconclusive. The one exception was mathematics which appeared to have a positive relationship with student performance.

Since Wayne and Young’s review, there have been several additional studies conducted on the relationship between teacher education level and student performance. Again results have been mixed. For example, Betts, Zau, and Rice (2003) found that students in San Diego’s Unified School District achieved slightly higher scores in math when their teacher held a master’s degree over a bachelor’s degree. Clotfelter, Ladd, and Vigdor (2006) on the other hand found that students from North Carolina’s school system did less well when their teachers had a master’s degree. Lastly, Harris and Sass (2007) found that “obtaining an advanced degree during one’s teaching career does not enhance productivity and may actually reduce productivity in high school math and middle school reading” (p.26).

Personal traits and characteristics. Research into the personal traits and characteristics of teachers and their impact on student performance is not a new line of inquiry. In fact, interest in this area has continued for much of the last century (Schalock, 1979). For example, early research in this area has shown a positive relationship between variables such as teacher adaptability, clarity, enthusiasm, task-oriented behavior, and variability of lesson approach, student opportunity to learn criterion material and student learning (Darling-Hammond, 1999). In addition, teachers who were able to structure their materials to ask higher order questions, use student ideas, and probe student comments had a greater impact on student learning than those who did not.

More recent research in this area has also looked into teacher expectations, efficacy, explanatory style, grit, life satisfaction, beliefs, attitudes, and values as they relate to teacher effectiveness (e.g., student achievement scores, administrator ratings, student gains). Sheftall (2000) looked at the relationship between teacher expectations and efficacy on student achievement and found that of the two, teacher expectations had a significant impact. In another study, Duckworth, Quinn, and Seligman (2009), looked at the relationship between teacher’s explanatory style, life satisfaction and grit, and teacher effectiveness. Optimistic explanatory style was defined as attributing bad things to specific and temporary events and good things to more global and long term events. Life satisfaction was defined as “...contentment with one’s

life situation” (p.541). Grit was defined as “...perseverance and passion for long-term goals” (p.541). Teacher effectiveness was defined as student gains. Results found that all three characteristics individually predicted student performance. However, when taken together only life satisfaction and grit remained significant. Finally, Metzger and Wu (2008) conducted a meta-analysis that looked at the relationship between the Gallup Teacher Perceiver Interview (TPI) (a well-known commercial instrument for selecting teachers based on their beliefs, attitudes, and values) and its impact on teacher quality. Teacher quality was defined in a variety of different ways to include: principal ratings, student ratings, classroom observations, student gain scores, and teacher attendance. Results found a modest relationship between TPI and some indicators of teacher quality to include administrative ratings and student ratings. Student gain scores were not found to be significantly correlated, although this result was based on only one study.

Interviews. One of the most commonly used methods of assessing job candidates for employment is the job interview (Latham, Saari, Pursell, & Champion, 1980; Levashina, Hartwell, Morgeson, & Campion, 2014; McDaniel, Whetzel, Schmidt, & Maurer, 1994;). According to Levashina et al. (2014), employment interviews have been used more than any other selection method in the last 100 years. In the realm of education this is no exception. In fact, it is the preferred approach for most school administrators when hiring teachers (Wise et al., 1987).

A job interview as defined by the literature is “a personally interactive process of one or more people asking questions orally to another person and evaluating the answers for the purpose of determining the qualifications of that person in order to make employment decisions” (Levashina et al., 2014, p.243). Job interviews can be either unstructured or structured. Unstructured interviews, as the name would suggest, tend to be informal in nature. There is no fixed format and content can often be unplanned or improvised (Winter, 1995). Winter (1995) indicated that the most common type of interview used in educational contexts is the unstructured interview. Structured interviews on the other hand, are more formal and organized. A structured interview is one that “...involves the establishment and deliberate application of predetermined rules for questions, observations, and evaluations” (Levashina et al., 2014, p.244).

Empirical research conducted on employment interviews is quite substantial in the business literature. According to Wise et al. (1987) researchers have been studying this area for over six decades now. In fact, over the last 30 years there have been 12 meta-analyses conducted on the topic (Levashina et al., 2014). Results from these efforts indicate that unstructured interviews are poor predictors of job performance (Latham et al., 1980; Wise et al., 1987; Winter, 1995; Dana, Dawes, & Peterson, 2013). Correlations between unstructured interviews and job performance typically range from .14 to .33 (Rogelberg, 2007). Conversely, structured interviews have performed much better with correlations between .35 and .57. According to Gimbert and Chesley (2009) structured interviews are more reliable in predicting job performance than unstructured ones. Levashina et al (2014) found that this is “one of the most consistent findings in the history of research on the employment interview...” (p.242).

In addition to being more predictive of job performance, structured interviews have been found to have incremental validity as well (Rogelberg, 2007). According to Levashina et al. (2014), structured interviews have been shown to provide incremental validity over personality

tests and cognitive ability tests. Rogelberg (2007) indicated that the criterion-related validity of a selection process can be raised by as much as 20% by adding a structured interview to the mix.

Work samples. The work sample test is another popular technique used by employers to select potential job candidates (Eurich, Krause, Cigularov, & Thornton, 2009). A work sample test is one “...in which the applicant performs a selected set of actual tasks that are physically and/or psychologically similar to those performed on the job” (Roth, Bobko, & McFarland, 2005, p. 1010). Consequently, work sample tests are seen as more of an approach than a single method. For example, work sample tests can include: hands-on tests, trainability tests, situational tests, job knowledge tests, and assessment center exercises (Callinan & Robertson, 2000).

Empirical research indicates that work sample tests have a consistently strong relationship with job performance (Callinan & Robertson, 2000; Darling-Hammond & Newton, 2013; Eurich et al. 2009; Roth et al., 2005; Thornton & Gibbons, 2009). In fact, Roth et al (2005) suggest both researchers and managers agree that it is among the most valid predictors of job performance available.

Empirical research in the education literature also shows a consistently strong correlation between work sample tests and job performance (Winter, 1995). For instance, several studies have found that a teacher’s ratings from their days as a student teacher (i.e., work sample test) were highly correlated with how well the teacher did during their first year of teaching (Schalock, 1979). Similarly, Boyd, Grossman, Lankford, Loeb and Wyckoff (2009) found that teachers who were given the opportunity in their teacher preparation programs to engage in teaching practices had greater student gains during their first year of teaching.

Instructor Preparation Methods

The following discussion of instructional method effectiveness is organized into two sections, the first focusing on the empirical research using instructor samples (teachers, trainers, facilitators) and the second using general student samples, other than instructors or teachers, illustrating instructional principles with which instructors should be familiar.

Instructor sample research. By far the majority of the empirical studies on instructional methods and techniques within instructor samples involve primary and secondary school teachers. A great deal of the literature on teacher effectiveness is based on survey and expert opinions, but the findings presented here will focus on the experimental research. Several meta-analyses and literature reviews of experimental studies are summarized.

Reviews concentrated on investigating optimal methods for preparing teachers, the role of teacher characteristics on learner outcomes, and the effects of both methods and characteristics on student learning, motivation and other outcomes. Unfortunately, confounding among experimental condition control, random participant assignment, subject characteristics and criteria was found in many studies. This confounding clouds estimation of the direct relationship between teacher training methods and student or other effectiveness outcomes (Blank, de las Alas, & Smith, 2008; Harris & Sass, 2007). For example, Yoon, Duncan, Lee, Scarloss and Shapley (2007) found that only nine of over 1,300 studies on teacher effects on

student outcomes met standards for conducting meta-analyses which limits the conclusions that can be drawn from such studies. That said, their results revealed that teachers, participating in an average of 49 hours of professional development, were able to boost student achievement scores by approximately 21 percentile points over the control group of teachers not participating in professional development activities. What qualifies as professional development has also been criticized as consisting of a “patchwork of opportunities—formal and informal, mandatory and voluntary, serendipitous and planned” (Wilson & Berne, 1999, p.174).

Though dated, Wade (1985) conducted an extensive meta-analysis of over 300 studies of teacher effectiveness and examined four specific in-service developmental techniques and their cumulative effect sizes on teacher and student related outcomes. Observation of effective instructors was found to have the highest cumulative effect size of (Cohen’s $d=0.81$), followed by micro teaching, (generally described as receiving feedback on one’s teaching by viewing a recording of one’s teaching and participating in review sessions with other teachers) ($d=0.78$), video/audio feedback ($d=0.64$) and practice ($d=0.55$) (note that the author did not clearly distinguish micro teaching from video/audio feedback). Each of these methods was more effective than the other methods reviewed, including lecture, discussion, games/simulations and guided field trips. These results indicate moderate to high levels of improvement with the experimental participant average performance exceeding 70% of the control group performance. While the findings related to the effectiveness of observing effective instruction and receiving feedback on instructing are important, the comparisons with technology assisted feedback are possibly less relevant today given the evolution of technology over the past thirty years and the expectations of students regarding the use of technology in instruction.

Recently, Pearce et al. (2012) conducted a meta-analysis of 18 studies examining the effectiveness of medical train-the-trainer (TTT) programs on participant knowledge, subsequent participant clinical behavior and patient outcomes. The training techniques included case studies and scenarios, lecture and other didactic presentations, video presentation, power point slides, group discussion, interactive methods, practical demonstrations and exercises, role plays, motivational and attitude change, problem-based learning and miscellaneous other methods. The findings support the use of TTT programs over no training with 13 of 18 studies showing significant effects on improving clinical behavior, clinician knowledge or better patient outcomes. Of interest was the finding that in one study the use of a CD-ROM training method was shown to be more effective than a live instructor on participant knowledge improvement.

Despite methodological limitations, Blank et al. (2008) found that about one-third of the studies they reviewed indicated significant improvements in teacher knowledge, changes in instructor classroom practices and student gain scores when examining teacher development programs that focus on content knowledge, coaching and mentoring, and other forms of peer collaboration. Cohen and Hill (1998), Kannapel and Clements (2005) and Wenglinsky (2002) found that professional development that is sustained, aligned with the curriculum, and focused on instruction is shown to positively influence student achievement in mathematics and science at both the elementary and high school levels. Yet many individual studies often indicate no significant differences in student outcomes despite concerted teacher development efforts. For example, Glazerman and Seifullah (2012) reported no differences in student learning despite their teachers’ participation in a Chicago Teacher Advancement Program.

Army instructors engage in a variety of professional development activities, though no formal program exists independent of the normal professional military education courses that are part of any military occupational specialty. The schoolhouses offer opportunities for instructors to engage in dialogue and to participate in workshops, seminars and other professional development activities within or outside the training and educational institutions. However, we found no specific evaluations of the effects of Army professional development on instructor effectiveness.

Induction and mentoring. Induction and mentoring of teachers has been investigated as a means for developing teacher skills and improving teacher retention rates. Teacher mentoring has been described as “a process to help novices develop teacher behaviors and strategies involving a nurturing relationship between a less experienced person and a more experienced person where the mentor provides guidance by serving as a role model and advisor” (Bigelow, 2002; Haney, 1997). This definition is similar to the Army’s definition (Department of Army, 2007) but the focus on developing teaching behaviors and strategies is more focused on job performance than in the Army. Induction is the term used for providing additional support, guidance and orientation to teachers in their first assignment or early in their careers. Smith and Ingersoll (2004) found that early stage teachers, provided with same subject mentors, and who participated with other teachers in planning and collaboration activities were about 30% less likely quit the profession than their counterparts with no induction or mentoring. No empirical results regarding the contribution of these programs to student outcomes could be found. The Army mentoring program is seemingly more focused on career development than improving instructional effectiveness, and therefore mentoring as the Army defines it is probably less likely to contribute to instructor skill and knowledge development.

Instructor certification. Teachers and various other instructors involved in high risk occupations (e.g., public safety, airline, medical) are typically certified in some formal way. Teacher certification in the U.S. is state-controlled and regulated by the Department of Education. Typically, education qualifications, knowledge of the subject in which the teacher will instruct and passing a nationally standardized certification test comprise teacher certification in public schools in the United States.

Despite the widespread use of teacher certification, the findings regarding effects on student outcomes are mixed. Goe (2007) found higher student math performance where teachers held subject level certification. In a longitudinal review of more than 15,000 teachers, Darling-Hammond, Hotzman, Gatlin, and Heilig, (2005) found after controlling for teacher experience, education degrees and student characteristics that certified teachers have higher student achievement levels than uncertified teachers. However, Cantrell, Fullerton, Kane, and Staiger (2008) found no significant difference between student math and language achievement scores between National Board for Professional Teaching Standards (NBPTS) certified and non-certified teachers. In addition, Constantine et al. (2009) documented that there were no significant differences in student achievement outcomes between traditionally certified and alternatively certified teachers. Goldhaber and Brewer (2000) found similar results when examining full versus emergency certification. Again the wide variety of contexts for teacher certification and performance likely inhibits conclusive results in terms of student achievement differences.

Army instructors are currently certified as described below in the section on Army Instructor Preparation and Certification Practices. Certainly the certification of Army instructors is a reasonable expectation for any instructional setting and program. Despite the lack of strong effects of instructor certification on student outcomes, the practice of certifying instructors continues to be a best practice in many different educational contexts and subject matter. While it would be ideal if instructor certification practices were shown to be linked to higher student achievement, learning transfer or even instructor satisfaction and motivation, the lack of findings does not discount the need to certify instructors in Army training programs. There are safety and classroom management issues involved in Army training that may not have been thoroughly evaluated in more traditional certification processes that support consistent instructor certification programs.

General population training effectiveness research. This section describes the results of research on various instructional methods and their effectiveness within general student populations, not specific to instructor development. General student population research on various learning methods is quite extensive and the assumption is that methods that are effective in preparing students will also be effective in preparing instructors.

Dunst, Trevette, and Hamby (2010) performed a meta-analysis of 58 experimental studies of instructional effectiveness coding study characteristics related to four different adult learning methods-and four study outcomes. The four learning methods included accelerated learning, coaching, guided design, and just-in-time training.

The four study outcomes included learner knowledge, skill acquisition, student attitudes and student self –efficacy beliefs. The results revealed average effect sizes across all outcomes of ($d=0.42$), with individual method effect sizes of coaching ($d=0.91$), just-in-time training ($d=0.52$), guided design ($d=0.49$) and accelerated learning ($d=0.05$). The results, generally, support the notion that the more actively the learners were involved in the learning activities (e.g., through planning, exercises and directing the learning), the greater the effects on the learning outcomes. These findings further lend credibility to the Army learning model and the goal of transforming military instructor development programs to encourage more facilitative and collaborative learning.

Lecture versus activity-based learning. Activity-based learning comprises the use of exercises, queries, problems, assignments and other similar activities performed individually or increasingly as part of small learning groups. Nearly all formal courses in primary, secondary, training and professional development use activity-based learning, multiple exercises, and small group interactions to develop knowledge comprehension, skill application and practice. Kalaian and Kasim (2013) conducted a meta-analysis of 193 studies that used collaborative, cooperative, problem- and inquiry based activities in combination with small group and teams in comparison with primarily lecture-based methods for each science, technology, engineering and mathematics (STEM) subjects. The mean effect sizes for activity-based learning in comparison with lectures were ($d=0.37$) on student achievement and ($d=0.31$) on promoting student interest in STEM and reduced college class withdrawal and failure by 7%.

Activity-based learning covers a large range of instructional methods and techniques, but many if not all of these techniques are routinely used in Army training and education programs.

We certainly would expect these methods to also be used when preparing instructors for formal and informal development of their Soldiers. More detail on various types of learning activities are described below.

Experiential learning. Experiential learning has been defined as “the process whereby knowledge is created through the transformation of experience. Knowledge results from the combination of grasping and transforming experience” (Kolb, 1984, p. 41). Experiential learning has now come to represent a broad swath of learning activities characterized by active student involvement and the application of existing personal knowledge and experiences into the educational environment (Bangs, 2011). An Experiential Learning Model has been used at the Army Command and General Staff College, and Meyers (2010) performed an assessment to reveal that students perceived the model to be effective in teaching critical thinking skills. Experiential learning for instructors occurs when they participate as instructors through practice, demonstration, and as they take knowledge of previous instruction and apply it to future contexts.

Problem-based learning. Problem-Based Learning (PBL) is an instructional model in which students are given a complex problem to solve that may not have a single correct answer (Hmelo-Silver, 2004). The teacher acts as a facilitator and guides the learning process through open-ended questioning, thus promoting self-directed learning and facilitating a sense of intrinsic motivation.

Instructional scaffolding. Scaffolding in the context of instruction refers to instructors and peers, computer-based tutors and avatars (Molenaar, Chiu, Sleegers, & van Boxtel, 2011) and other materials (Puntambekar & Kolodner, 2005) providing support and structure for student learning activities. Materials may include advanced organizers, cue cards, concept and mind maps, examples, handouts, and other prompts. The evidence for the effectiveness of scaffolding has been demonstrated in a variety of contexts, including with literacy instructors (Pressley, et al., 2001), improving interpretation and understanding of clinical trial research (Dawn, Dominguez, Troutman, Bond, & Cone, 2011), and in support of online learning of critical inquiry skills (Bai, 2012).

Situated and authentic learning. Situated and authentic learning practices are based upon the notion that learning occurs and is influenced by the culture, context, and activities in which the performance takes place (Lave, 1988). Situated learning stipulates that “knowledge be presented in authentic contexts (settings and application that would normally involve that knowledge) and learners to participate within a community of practice” (Naismith, Lonsdale, Vavoula, & Sharples, 2004, p. 13). On-the-job education and training are examples of situated learning. For Army instructors, the notion of situated and authentic learning would suggest that instructors are best developed in the same environments in which they operate, namely the classroom, range, simulator or other environments where learning takes place. It would also follow that Army instructors should be prepared using the same methods by which they will teach their students.

Competency-based education/learning. Competency-based education or learning is predicated on students being able to demonstrate some level of proficiency with competencies, defined in many ways but generally representing a set of knowledge, skill, ability or behaviors

and actions that are related to success or effectiveness on the job or beyond the classroom (Shandler, 2000).

Our recommendations for preparing instructors will be based upon our definition of an effective instructor and the knowledge areas, skills, abilities and other constructs that support effective instructional behaviors. This is consistent with competency-based instructional principles.

Self-directed learning and self-regulated learning. Knowles (1975) defines self-directed learning (SDL) as “a process in which individuals take the initiative, with or without the help from others, in diagnosing their learning needs, formulating goals, identifying human and material resources, choosing and implementing appropriate learning strategies, and evaluating learning outcomes” (p. 18). Murad, Coto-Yglesias, Varkey, Prokop, and Murad (2010) conducted a meta-analysis of 59 studies that revealed that the use of SDL was associated with a moderate improvement in knowledge-based outcomes compared with didactic instruction, but there were no significant differences between the two with respect to skill and attitude-based outcomes. The study also found SDL to be more effective when learners were involved in identifying their own learning resources.

Self-regulation learning (SRL) theory posits that the learner manages affective, cognitive, and behavioral processes throughout a learning experience to reach desired goals. Using SDL and SRL techniques may represent an opportunity to support Army instructor ongoing development beyond current institutional courses as individual instructors create their own learning path and developmental activities.

Collaborative and cooperative learning. Collaborative and cooperative learning (the terms are used interchangeably) involve interactions between learners to improve the learning experience. Cooperative learning has been defined as working together with another person or group to accomplish shared goals (Lefrancois, 1999, p. 539).

This certainly seems to be a very fertile area for research and we recommend the Army investigate methods by which instructors can use collaboration and cooperation to build their skills and better prepare for facilitating learner-centered environments. This would be particularly true for instructors following formal institutional courses upon assignment to their institutional programs.

Blended learning. Blended, hybrid, and e-learning, which combine online with traditional classroom learning, refer to an evolving set of definitions combining: face-to-face and online instruction, media and technologies, and pedagogical methodologies (Sharma, 2010).

Improved learning outcomes under blended learning conditions have been demonstrated by Boyle, Bradley, Chalk, Jones and Pickard (2003) and Dowling, Godfrey, and Gyles (2003). Starenko, Vignare, and Humbert (2007) described the results of an earlier study showing student satisfaction increased under blended learning conditions.

Computer-based learning environments. These methods for learning include simulations, games and social collaboration. Training simulations and simulators have been found to be effective and efficient in a wide variety of skill based domains, including flight skills (Hays,

Jacobs, Prince, & Salas, 1992), surgical procedures (Hague & Srinivasan, 2006), and marksmanship (White, Carson, & Wilbourn, 1991). When examining specific features of computer-based instruction, Ma, Adesope, Nesbit and Liu (2014) found that intelligent tutoring built into computer-based learning programs had significant effects ($d=0.42$) over large group instruction but showed no advantages over small group instruction.

Learning games represent a unique type of simulation for learning. Sitzmann's (2011) meta-analysis of computer-based simulation games revealed that while self-efficacy, procedural and declarative knowledge and retention may be increased, the methods may not offer advantages over traditional instructional methods that include engaging techniques.

It seems unlikely that, absent a significant development effort, games or simulations would represent a viable option for preparing Army instructors other than those that would be used with their students and those simulators included in existing weapons and other technical courses.

Apprenticeship. Apprenticeship is a form of education where a master craftsperson provides direct instruction to a student or an apprentice by passing on the skills and knowledge of the particular occupation (Brewer, 2011). While the Army does not have a formal instructor apprenticeship program, aspects of apprenticeship, including receiving critiques on their instruction, mentoring, coaching and senior instructors providing feedback and guidance to junior instructors are included in most, if not all, Army instructor development programs.

Army instructor preparation and certification practices. Non-commissioned officer education system (NCOES) instructor preparation and certification guidance is provided in TRADOC Regulation 600-21 (2013a). NCOES instructor development includes courses in basic and advanced instructional concepts, small group instruction, systems approach to training, test construction and development and evaluating instructors. Both common (Staff and Faculty Common Training) and local (Staff and Faculty Local Curriculum) courses are available. TR 600-21 provides recommended instructor training (e.g., courses, modules, workshops, guidelines and other materials) corresponding to competency training at three levels of instructor recognition (instructor, senior instructor and master instructor) for nineteen instructor competencies.

Instructor certification and recertification requirements are described in TRADOC Regulation 350-70, and include foundational course completion, serving as instructor/facilitator of one or two lessons of the course they will instruct under evaluation by certified instructors in that course, and demonstrated subject matter expertise and proficiency in the instructional techniques for delivering that course under a certified instructor for a period of 30 days or less as determined by the institution. Additional Army instructor certification information can be found in section 4-2 of TR 350-70.

Other military instructor development and certification practices. U.S. Navy, Air Force and Marine Corps instructor development is similar to the Army in that instructors have established preparation and certification requirements at multiple levels of instructor competence and experience levels. Air Force Air Education Training Command Instruction 36-2202 (Department of Air Force, 2012) indicates instructor development programs consist of various

basic and intermediate courses on teaching methodologies, questioning techniques, academic counseling, core values and professional relationships. Courses are taught within Faculty Development units which oversee instructor technical training and development of instructors within the Community College of the Air Force (CCAF). The CCAF teaches collegiate level courses leading to degrees. Instructors contributing to course development must also take instructional systems development and technical writing courses. Teaching internships (180 hours) are also required within all Air Force courses, with some exceptions. Instructors are also certified by more experienced instructors. Air Force instructors carry specific occupational designation and categories of instructors include first-tour, returning, in-service, supervisory, and master instructor.

Navy instructor preparation, qualification, certification and evaluation program information is contained in Naval Education and Training Center Instruction 1500.5 (series). In 2012, the Navy revised the Navy Instructor Training Course (15 instructional days), to replace the Journeyman Instructor Course, adding 40 hours of contact time to that course. Additional focus on lifelong learning and continual Sailor development were included in the curriculum upgrade. An instructor Navy Enlisted Classification (NEC) is conferred upon graduation and certification of instructor skills through evaluation by senior instructors, and the Sailor continues to keep the NEC for future instructor selection decisions.

Marine Corps Order 1553.2b (Department of Navy, 2011) covers preparation and certification requirements for Marine instructors. Similar to the other services, Marine Corps instructor development includes basic instructor courses taught at formal learning centers and demonstration of skills through observation of teaching. Marine Corps instructor development has recently focused on small group facilitation and cognitive readiness which is described in more depth in the following section.

Recent military instructional and certification research. Recent interest in improving Army instructional methods cover problem-based learning (Cianciolo et al., 2011), cognitive readiness and other instructor professionalization activities (Schatz et al., 2012), adaptive training (Schaefer & Dyer, 2012), peer-to-peer learning (Cooper, Leibrecht, & Lickteig, 2011), and the use of games in instruction (Beal, Wright & Topaz, 2009). While none of these efforts included empirical evaluations of the effectiveness of these methods on instructor or student outcomes, they represent the wide range of possible methods being investigated for improving instructor training and development.

The Marine Corps has recently implemented two instructor development initiatives focusing on improvement of instructor and student cognitive readiness, defined as “the mental ability necessary to survive in a complex and unpredictable combat environment,” (Morrison & Fletcher, 2002). The first initiative seeks to enhance small unit decision making principles including facilitative strategies toward specific problem solving that includes focus on self-awareness, attentional control, meta-cognition, problem solving and sensemaking (Schatz et al., 2012). This initiative has been implemented through instructor development seminars, course reviews and revisions, and other materials such as handbooks and instructor guides. The second initiative targets improving instructor professionalism through the development of instructional competencies, expanding the tools available to facilitate learning, and focusing on direct,

indirect, interactive, independent and experiential learning tactics (Schatz et al., 2012). To date, we could find no empirical findings regarding the effectiveness of these programs.

Instructor Evaluation Methods

As previously noted, education literature suggests there is a lack of clear consensus on what an effective teacher is and what an effective teacher does. This can serve as a barrier for assessment and evaluation, and some in the education field argue that without a definition of effective teaching, teaching cannot be evaluated. Yet others more plainly believe teaching is too complex and subjective to be evaluated in the first place (Seldin, 2006). Therefore, it is not surprising that across settings there appears to be no generally agreed-upon method for evaluating teacher effectiveness. It has further been noted that the methods used for evaluating teachers have changed as definitions and beliefs about what is important to measure have evolved (Goe, Belle, & Little, 2008).

The existing literature on instructor evaluation is rich with suggested frameworks and guidelines for the design of assessment and evaluation systems. However, while regular and consistent feedback on classroom instruction can be a powerful way to improve teacher effectiveness, studies have found that teacher assessments and evaluations are not typically seen as useful tools (Oliva, Mathers, & Laine, 2009). Current systems that are used to assess, evaluate and support teachers too often fail to improve teacher practice and enhance outcomes such as student growth and learning. The National Education Association (2010) notes that to be effective, evaluation systems require an infrastructure of key components such as trained classroom observers, carefully designed assessment instruments, and the ability to provide teachers with constructive, actionable feedback for improvement.

This review begins with operational definitions for key terms relevant to the measurement of instructor effectiveness. While existing literature commonly uses the terms assessment and evaluation interchangeably, these practices differ in important ways. The National Education Association (2010) offers useful distinctions between teacher assessment and teacher evaluation in a paper on transforming education systems to support effective teaching. Specifically, *assessments* refer to practices intended for instructor growth and improvement that are diagnostic in nature and occur on a continuous basis. Assessments are individualized and generally collegial to encourage self-reflection on the part of the instructor. Thus, instructor assessments are *formative* in that the feedback received has meaning to the individual and is useful for personal improvement. In comparison, *evaluations* refer to standards-based measures that are judgmental and hierarchical in nature and occur on a periodic or scheduled basis. Evaluation often involves a rubric of criterion-referenced measures to establish *summative* judgments about an instructor's effectiveness. Wiliam (2006) acknowledges that assessments and evaluations are not distinguished by the format of a measure but rather how the information from a measure is used. The same measure may be employed for both formative and summative purposes. The current review uses the term evaluation, for the sake of clarity.

Authors have argued that formative and summative purposes of measures have become confused in practice, and that as a consequence, assessment often fails to serve a truly formative purpose (Harlen & James, 1997). Seldin (2006) notes that colleges and universities evaluate

faculty members for two primary reasons: to improve their performance, and to provide rational and equitable basis for personnel decisions. Ideally, faculty evaluation would be conducted separately for the purpose of improving teaching and gathering information for personnel decisions. However, time and financial constraints often lead institutions to conduct them simultaneously by integrating them into a single questionnaire rating dimensions that serve both purposes.

Goe et al. (2008) also distinguish high-stakes from low-stakes measurement of teacher effectiveness. For example, an informal classroom observation by a supervisor that does not carry serious consequences and is meant to provide formative feedback to improve teaching is considered *low-stakes*. In contrast, a formal evaluation that carries substantial consequences (e.g., conducted to gather information for specific decision-making processes) is considered *high-stakes* and summative. The authors note that considering the intent of teacher evaluation (i.e., whether it is high-stakes or low-stakes, formative or summative) has strong implications for choosing a measure that will provide valid results.

Teacher quality and teacher effectiveness. In the review by Goe et al. (2008), the authors note the No Child Left Behind (NCLB) Act mandates that all teachers should be highly qualified, though clearly being “highly qualified” (i.e., having the necessary qualifications and certifications) does not necessarily predict highly effective teaching that improves student learning. Too often teacher effectiveness is defined as the ability to produce gains in student achievement scores, though this concept is far too narrow.

Goe et al. (2008) propose useful distinctions between three different but related angles for evaluating teacher effectiveness. First, *teacher inputs* represent what a teacher brings to the role, and include the teacher’s background, beliefs, expectations, experience, pedagogical and content knowledge, certification and licensure and educational attainment. These elements generally reflect KSAOs by which teachers are identified or selected. Second are *teacher processes*, which refer to the interaction that occurs between teachers and students but also a teacher’s professional activities outside of the classroom within the larger institution and community. These elements generally reflect work behaviors that teachers demonstrate as part of their job. Finally, *teacher outputs* represent the results of classroom processes such as impacts on student achievement, behavior, engagement, attitudes and social-emotional well-being. The authors propose that the teacher inputs be referred to as *teacher quality*, while the teacher outputs be referred to as *teacher effectiveness*.

Thus, assuring teacher effectiveness (i.e., positive teacher outputs) begins with the use of sound methods for teacher selection and preparation. That is to say, education systems that begin with a focus on teacher quality (i.e., positive teacher inputs) can ensure every teacher demonstrates subject-area knowledge, pedagogical knowledge, and professional teaching ability. Importantly, instructor selection practices support teaching effectiveness when the criteria for hiring or identifying teachers align with the criteria used for evaluating teachers (National Education Association, 2010). Effective selection practices can account for the level of preparation and experience teachers bring to the role. From there, ongoing (cyclical) methods of assessment and preparation can serve to continuously measure and improve teacher effectiveness.

This review addresses assessment and evaluation of instructors that originates with the instructor's interaction with students in the learning environment. Said another way, the focus is on measurement of effective instruction at the point where the instructor brings to bear the instructional tools and strategies to positively impact student outcomes. This does not include assessment or evaluation practices used during earlier institutional or instructional processes (i.e., identification and preparation). This is an important distinction, as methods of instructor (or candidate) selection and preparation inherently include elements of assessment whereby relevant factors of teacher quality are considered and/or evaluated.

Instructor evaluation in industry. In industry, instructor evaluation is often a component of training evaluation, and Kirkpatrick's Four-Level Evaluation Model (Kirkpatrick, 1994) is perhaps the most prominent and widely-used framework for evaluating training courses and programs (Hilbert, Preskill & Russ-Eft, 1997; Hoole & Martineau, 2014). Kirkpatrick observed that most outcomes of training and development could be categorized as reactions, learning, behavior and results (1959a, b; 1960 a, b). These four levels individually provide a unique lens into the success of training courses and programs. At *Level 1* (Reaction), measures capture outcomes about how people think and feel about the training, including the instructor, training topics, content, presentation, and how relevant the training is to the learner's work. Reaction data are often obtained through trainees providing ratings on a post-training survey. *Level 2* (Learning) consists of cognitive measures to determine how much learners' knowledge has increased due to the training. Measurement typically involves a test of the learners' content knowledge. At *Level 3* (Behavior), measures evaluate how much learners have changed their behavior based on the training. In practice, this includes examining the degree with which learners use a new skill on the job and their effectiveness in doing so. Finally, *Level 4* (Results) involves an analysis of the effects of training on outcomes that are positive for the organization. Here, measures assess the impact of improved individual performance on organizational indicators (outcomes) such as increased productivity among workers who received the training. Notably, Level 4 is often difficult to evaluate as organizations have limited resources with which to evaluate training effectiveness. In addition to Kirkpatrick's four levels, Phillips (1983) adds return on investment as *Level 5* of training evaluation, whereby results data are converted to monetary values to compare the benefits of a training program to its costs.

The relevance of the Kirkpatrick Four-Level Evaluation Model (1994) to this effort's definition of an effective instructor is as follows. As an example, in Level 1 (Reaction), learners may provide feedback on training elements such as instructor teaching style, the instructional strategies used, methods of presentation, and pace (all reflective of the instructor), as well as other aspects of the training experience. In this way, learners are an important source of data on instructor effectiveness, particularly for elements such as an instructor's capacity to use appropriate strategies and techniques and to demonstrate empathy and a personal capability to tailor instruction based on individual needs. The degree with which learners retain knowledge (Level 2, Learning) and demonstrate new knowledge or skills on the job (Level 3, Behavior) can serve as outcomes-based measures of instructor effectiveness. However, while evaluation at Levels 2 and 3 are useful in that they may point to an instructor's ability to create positive learner outcomes, the evaluation is not directly inclusive of the instructor's application of appropriate strategies and techniques to achieve those outcomes. Measurement of outcomes can provide an

indication of successful instruction, but outcomes alone do not point specifically at what an instructor does in the teaching context to influence the result in learning or behavior. Level 4 (Results) (and arguably, Level 5, Return on investment) evaluates outcomes of training even further removed and less attributable to the influence of the instructor.

At a level more specific to instructor evaluation, guidance by the American Society for Training and Development (ASTD, now Association for Talent Development) recommends general practices for measuring instructor effectiveness. These include the use of multi-dimensional instructor competencies, collecting data from multiple sources, employing multiple methods, avoiding a competitive mentality (e.g., unspoken agenda), and distinguishing meaningful from meaningless data (Conway & Cassidy, 2001). Meaningful and useful evaluation data are collected at multiple points in time, and random variation in instructor effectiveness should be expected. Four key sources of data for measuring trainer effectiveness are trainees, fellow trainers, training management and trainers themselves. Together, these sources offer measurement analogous to 360-degree assessments, as all sources offer performance feedback from a unique vantage point of trainer effectiveness (i.e., self, peers, supervisors, and training recipients). Using a multi-source approach for instructor evaluation increases the ability to capture information on instructor capabilities such as applying the appropriate instructional tools, creating positive student outcomes, and demonstrating empathy to tailor instruction. However, the effectiveness in measuring these considerations also relies on the fidelity of the measure used and its application.

Instructor evaluation in the Army and other services. Within the Army, TRADOC schools and centers are responsible for comprehensively assessing the performance of their instructors on a regular basis. Army regulations indicate that this is done by observing an instructor's ability to follow lesson plans, teach to the standard, use instructional media properly and detect (and respond to) student needs appropriately (U.S. Army Training and Doctrine Command, 2011). The Army regulation that provides guidance on assessment and evaluation of NCO instructors is TRADOC Regulation 600-21, *Noncommissioned Officer Education System Instructor Development and Recognition Program* (2013a).

TRADOC Regulation 600-21. As previously referenced in discussions on instructor selection and preparation, TRADOC Regulation 600-21 governs implementation of the Army's Noncommissioned Officer Education System (NCOES) instructor development and recognition program (IDRP). Chapter 4, Policies and Procedures, describes instructor competencies, training and education, assessments, and recognition requirements. These practices within the IDRP encompass competency-based requirements for instructors. The list of 19 instructor competencies along with performance outcomes for each level of instructor recognition are presented in Appendix D of this regulation. A majority of the competencies (18 of 19) are copyrighted by the International Board of Standards for Training, Performance and Instruction.

The instructor assessment instruments that are also presented in the appendices of this regulation are summarized in Table 6.

Table 6

TR 600-21 Instructor Assessment Materials.

Instrument or Tool	Description
Instructor Competency and Outcomes Matrix (Appendix D)	Includes descriptions of 19 instructor competencies and outcomes specified at three tiered levels of instructor recognition: Instructor, Senior Instructor, and Master Instructor.
Instructor Self-Assessment (Appendix G)	Serves as an informal tool to assess instructor strengths and weaknesses in competencies and to guide development activities for self-improvement. Results of the self-assessment are to be shared with one's supervisor and compared with results of the most recent evaluation. Instructors rate themselves on a 4-point scale from strongly disagree to strongly agree (representing Not performed; Incorrectly or Incompletely performed; Satisfactory; Proficiently).
NCOES Instructor Observation Rubric (Appendix H)	Tool is used to evaluate instructor performance in teaching. The results are formative and used to update the instructor's self-development plan and to determine successful progression through instructor levels. <ul style="list-style-type: none">• Section 1 of this rubric covers the classroom environment (Ratings of Go, No Go, and N/A).• Section 2 covers 16 dimensions aligned with the instructor competencies, rated on a scale of Unacceptable, Developing, Accomplished, and Exemplary, with qualitative feedback on the dimensions followed by a written summative evaluation.
Course/Lesson Design Checklist (Appendix I)	Contains evidenced-based instructional design strategies used for the design and redesign of lessons. The checklist captures whether an instructor meets guidelines for instructional media selection; evaluating course/lesson introductions; evaluating conceptual, process and procedural knowledge design; practice feedback and assessment design; and evaluating course/lesson summaries. A qualified evaluator rates whether the 45 elements/sub-elements of the lesson are met (GO/NO GO) and may provide qualitative remarks for each.

Additionally, this regulation includes a competency assessment matrix that aligns each of the 19 competencies (for each of the three tiered levels of instructor recognition) with the appropriate

methods or tools for assessment (presented in Appendix F). For example, the instructor competency “demonstrate effective presentation skills” at the instructor and senior instructor level is assessed using student questionnaires, items on the instructor observation rubric, and items on the instructor self-assessment.

The Air Force, Navy, Marine Corps and Coast Guard generally utilize similar approaches to assess instructor performance and to ensure the quality of training delivery. Often, evaluation practices are integrated within a Service’s instructor certification and professional development program. Programs for instructor preparation, certification and evaluation also tend to be tiered by levels of proficiency and experience. Common evaluation methods include classroom observations by trained evaluators, checklists with predetermined criteria (e.g., instructor competencies, procedural steps to teaching), self-assessments, and quantitative and qualitative feedback provided to the rated instructor as part of a debrief, with action plan for improvement (when necessary). The various observation checklists used during classroom observations tend to capture ratings on the presence and effectiveness of instructor teaching behaviors (e.g., using training aids effectively, maintaining control of class) and other steps or processes used during the course of instruction (e.g., provided lesson overview, safety brief, recapped key points). The following summaries highlight current practices within each of the other Services.

U.S. Air Force. Air Education Training Command (AETC) Instruction 36-2202 (2012) provides procedural guidance and responsibilities for planning, conducting and documenting training and evaluation for instructors. The intent of the instructor evaluation process aims to achieve both formative and summative objectives, which includes evaluating the quality of instructor performance and providing constructive feedback to improve training delivery. Evaluation is done to ensure instructors apply effective teaching methods and techniques and to guarantee overall consistent training delivery. Evaluations are tiered across the following types:

- Initial qualification of a student instructor’s mastery of teaching methods and techniques;
- Scheduled evaluation of an instructor’s ability to teach without assistance;
- Follow-up evaluation for instructors rated as needing improvement;
- No-notice evaluations performed outside the typical schedule; and
- Master instructor evaluations to determine qualification for the master instructor award.

Trained evaluators, designated by the squadron commander, utilize a one-page evaluation checklist consisting of five instructor proficiencies assessed by 28 items. Ratings are made on a scale consisting of Outstanding, Excellent, Satisfactory, Needs Improvement, and Not Applicable. While the form itself offers no space for qualitative notes or observations, the guidance states that evaluation feedback is to be provided to the instructor in a constructive manner with specific recommendations for improvement, when necessary. Any remediation for below satisfactory performance is handled by the instructor’s supervisor, who facilitates appropriate actions and/or additional training. Upon initial qualification at each instructional level, instructors receive 30, 60 and 90 day evaluations, and then annual evaluations beyond that point.

U.S. Navy. Guidance on the U.S. Navy’s instructor preparation, qualification,

certification and evaluation program is described in Naval Education and Training Center Instruction 1500.5 (2010). Upon initial qualification, instructors are evaluated through formal instructor performance evaluations and staff/student survey feedback both to assess instructor performance and identify opportunities for training improvement (i.e., formative and summative purposes). Minimally, instructors are evaluated semi-annually, while those with a master training specialist qualification are evaluated annually. Trained evaluators observe instructors in the classroom and utilize an instructor evaluation checklist consisting of 55 items categorized within five dimensions and having a rating scale of Satisfactory, Needs improvement, Unsatisfactory and Not observed. The form also includes a summative overall grade for the instructor's performance as well as summative qualitative remarks about strengths and areas requiring improvement. Instructors failing to maintain the original screening/selection requirements as well as those receiving unsatisfactory evaluations are disqualified, though re-certification may occur when deficiencies have been corrected or standards are met.

U.S. Marine Corps. The U.S. Marine Corps' guidance on instructor preparation and certification requirements is outlined in Marine Corps Order 1553.2b (Department of Navy, 2011). Formal and informal methods of instructor assessment begin following certification. A component of the Marine Corps staff and faculty development program consists of formative measures of assessment and development for new instructors. Examples include observations and reviews of new instructors' teaching ranging from non-threatening and professional feedback by peers to rigorous evaluation and informal certification by staff (typical of the "murder boards"), as well as videotaping of presentations or discussions for self-analysis.

More broadly, Marine Corps Formal Learning Centers (FLC) conduct comprehensive course evaluations on an ongoing basis by collecting data from multiple sources including students (instructional rating forms, end-of-course critiques), graduates (post graduate surveys), supervisors of recent graduates (surveys), and course instructors (after instruction report). Instructors and their learning environments are assessed during classroom observations, whereby an observer uses a series of checklists to rate instructor performance, lesson quality, the learning environment, and safety considerations.

- The instructor evaluation checklist covers 11 dimensions assessed by 49 items on a scale consisting of Yes, No or Needs Improvement. Qualitative comments are recorded in the margins of the form and the checklist ends with a summative rating and remarks section. The form includes an instructor improvement plan as part of the debrief process between the instructor and the observer.
- Following a lesson, student reaction to instruction is captured on an instructional rating form that includes quantitative items that assess the instructor's knowledge depth, communication skills, and use of instructional techniques, along with students' general reaction to the lesson content and learning environment. Then upon course completion, an end-of-course critique form captures student reaction to the course and includes items on instructor methods, knowledge and preparation, and professionalism.

U.S. Coast Guard. The Standard Operating Procedures (SOP) for the Coast Guard's Training System (2011) provides qualification requirements for the five professional training billets in the Coast Guard Training System, including instructors, master training specialists,

instructional designers, certified performance technologists, and training managers. Coast Guard instructors must qualify within 6 months of reporting to their assignment. While the initial stages of qualification involve completion of the instructor development course and meeting personnel qualification standards, instructors must also complete three classroom presentations and receive satisfactory evaluations. Classroom observers complete an instructor feedback form that measures 14 competencies on a quantitative scale. The form has an additional qualitative comment section for remarks on weaknesses and strengths by competency and is meant to serve as a basis for formative improvement strategies.

Instructor evaluation in academia. This section reviews methods and practices for evaluating instructors in post-secondary education and in public K-12 education.

Post-secondary education instructors. The evaluation of institutional faculty differs slightly from evaluation of instructors in traditional teaching roles, though common methods are applied. Paulsen (2002) notes that at the post-secondary level, actual teaching competes with other faculty activities such as research. Increasingly, institutional faculty face expectations to create student-centered classroom learning environments, focusing on active learning, the use of techniques for classroom assessment and research, and developing pedagogical content knowledge. However, faculty rewards are rarely linked to these types of teaching innovations.

At a broad level, best practices for applying effective faculty evaluation in institutions include clarifying expectations of and by faculty, identifying the nature and sources of data used for evaluation, and clarifying the purposes and uses of evaluation data (Cashin, 1996; Paulsen, 2002). However, to identify teaching responsibilities, the question of what constitutes effective teaching within an institution must be addressed. Like other settings, there is no universally accepted definition of effective college teaching.

A review by Canale, Herdklotz, and Wild (2012) examined the instructor evaluation practices at thirty universities and found that most institutions did not readily specify an institution-wide program of teaching evaluation. Rather, many institutions had policies requiring teacher evaluation (as stated in faculty handbooks, promotion and tenure guidelines, and other human resources policy documentation) but did not explicitly state methods or practices for evaluation. Further, it was found that teaching evaluation is often administered at the department level and by other faculty using resources disseminated via faculty development departments and/or centers for teaching excellence. The study reported that the most common teaching evaluations at benchmark institutions included peer evaluations (colleague and senior faculty), classroom observations (third party), small group instructional diagnosis, and use of teaching portfolios.

Primary and secondary education instructors. Authors Oliva, Mathers, and Laine (2009) posit that effective teachers are the greatest school-based contributors to improved student outcomes, and thus education systems must provide meaningful ongoing (formative) and summative feedback to teachers. A study by Brandt, Mathers, Oliva, Brown-Sims, and Hess (2007) examined common evaluation policy components at 140 schools across seven mid-western states. The authors identified several gaps between teacher evaluation best practices compared to current methods used in schools. For example, administrators or principals are the

most common evaluators of teachers, though a best practice is to use multiple evaluators such as teacher mentors or peers with a common instructional background. It was found that evaluators are rarely required by policy to be trained, though a lack of training introduces potential bias to the evaluation. To be authentic, evaluators must understand the evaluation rubric and characteristics and behaviors that the evaluation is intended to measure. Evaluations for non-tenured teachers tend to occur twice per year, while tenured teachers are evaluated every two to five years. Oliva et al. (2009) suggest that infrequent evaluations result in missed opportunities for formative feedback and improvement, but note that more research is needed to determine optimal frequency of evaluations for both non-tenured and tenured teachers. A general finding regarding communication was that district policies do not always require that teachers be informed about the criteria, process and implications of evaluations they receive. Authors have recommended that systematic communication occur with teachers before, during and after the evaluation process (Darling-Hammond, Wise, & Pease, 1983; Stronge, 1997).

Regarding evaluation instruments and measures, authors Goe, Bell, and Little (2008) conducted a research synthesis of approaches for evaluating teacher effectiveness. The review considered recent, empirical research studies from peer-reviewed journals that addressed the K-12 student population. The resulting synthesis examined 120 studies, focusing primarily on instruments and measures that more directly assess the processes and activities that occur during instruction and the products that are created inside classrooms. The study revealed the most widely used methods for evaluating teacher effectiveness include classroom observations and value-added models. Other methods for evaluating teacher effectiveness include portfolios, analysis of artifacts, teacher self-reports, analysis of student work, student ratings, and other reports such as documenting teacher's positive contributions to the school and teacher's leadership and mentoring. Aside from these methods, student achievement scores continue to be the focus of measuring teaching effectiveness. Several of these common teacher evaluation instruments were also identified within the aforementioned study by Brandt et al. (2007), including classroom observations; lesson plans; portfolio assessment, student work samples, and other instructional artifacts; self-assessments; and student achievement data.

More recently, a three year study titled the Measures of Effective Teaching (MET) project investigated better ways to identify and develop effective teaching. The project, funded by the Bill and Melinda Gates Foundation (2013), examined three commonly used measures of effective teaching in seven school districts nationwide. Measures included classroom observation instruments, student perception surveys, and student achievement gains. General findings of this research indicated that effective teaching can be reliably measured, that the use of multiple measures produces more consistent ratings than student achievement measures alone, and that use of more than one observer increases reliability significantly more than having a single evaluator conduct more than one observation.

A general finding of this review is that school administrators favor the use of student growth and/or scores as measures of effective teaching, as opposed to measures that provide more specific information on teacher practice in the classroom (e.g., observations, peer evaluations, self-assessments). Proponents of achievement metrics tend to lean toward rationalizations for improving school efficiency and making students better being the drivers for the market. This often means that things like teacher quality and student achievement tend not to

be clearly articulated and the metrics for determining these values are often questionable and result in policies with teachers being evaluated and compensated based on students' scores on standardized tests. However, standardized tests (i.e., student achievement) have been found to be an unreliable metric at the classroom level of analysis.

Research on methods for instructor evaluation. This section reviews the most prominent methods used for instructor assessment and evaluation. While a great deal of the literature is based upon expert opinion for best practice, these summaries cite relevant empirical research where available. As with instructor preparation, a majority of the studies on methods for instructor assessment and evaluation examine primary and secondary school teachers, followed by research within post-secondary academic settings. The methods reviewed include classroom observations; student achievement and value-added modeling; student evaluation of teaching; self-assessments; and portfolios.

Classroom observation. Observations are widely used to measure classroom processes such as teacher instructional practices, holistic aspects of instruction, and interactions between teachers and students. Observations are often used to measure teacher practice or behavior against some standard of effective teaching. Thus, it is important to carefully identify and define *what* is to be observed before conducting an observation (Berry et al., 2012; Goe et al., 2008; Oliva et al., 2009). Classroom observations are generally conducted by administrators, principals, supervisors, peers or other third party observers, depending on the instructional setting. In relation to this effort's definition of an effective instructor, classroom observations are useful for measuring instructor effectiveness at applying appropriate instructional tools (strategies and techniques), and may provide indication of an instructor's ability to demonstrate empathy and a personal capability to tailor instruction based on individual differences.

In primary and secondary schools, classroom observations by principals or vice-principals are one of the most common forms of evaluation (Brandt et al., 2007). Principals are the most knowledgeable about their schools, but are also more likely to compare teachers to each other. Principal observations occur formally (e.g., scheduled, using a validated instrument) for formative or summative purposes, or can consist of an informal drop-in to obtain a quick impression of how a teacher is performing in the classroom (Goe et al., 2008).

Peer observations have become increasingly common in university settings (Berry et al., 2012). In such settings, peer instructors have the requisite domain knowledge and expertise required for meaningful assessment and evaluation of teaching on aspects such as content mastery, course goals, course organization and materials (Paulsen, 2002). Thus, peer review of teaching brings content-based contextually to the evaluation of teaching. However, faculty peers are much more accustomed to reviewing one another's research as opposed to methods and practices for teaching. Authors note that when observations are conducted by peers, a collaborative approach is favored. This involves identifying what specifically will be observed, who will conduct the observation, and how the process will work (Bell, 2002; Berry et al., 2012). Peer observations are an effective method for formative assessment, as peers can structure constructive criticism, share best practices, and engage in peer-to-peer mentoring (Ammons & Lane, 2012).

In general, research has found positive relationships between observation scores and important outcome measures such as student achievement (Gallagher, 2004; Kimball, White, Milanowski & Borman, 2004). A study by Kane et al. (2012) investigated five different approaches (instruments) for classroom observation with a sample of over 1,300 teachers. Each instrument was designed to focus the observer's attention on specific aspects of teaching practice and to establish common evidentiary standards for each level of practice. Importantly, the instruments were not checklists (i.e., focusing on easy to measure but trivial aspects of practice) but rather required training and judgment on the part of the observer. The study objective was to compare instruments using two criteria: instrument reliability and association with student outcomes. The findings demonstrated that all five observation instruments were positively associated with student achievement gains (i.e., positive student outcomes). However, reliably characterizing a teacher's practice required averaging scores over multiple observations. Single observations produced reliabilities ranging from 0.14 to 0.37, while reliabilities around 0.65 were achieved only by scoring four different lessons. Combining observation scores with evidence of student achievement gains and student feedback improved predictive power and reliability, which is support for the use of multiple measures in evaluating teacher effectiveness.

Other recent research by Ho and Kane (2013) examined the accuracy and reliability of school personnel in performing classroom observations. The findings reinforced the need for more than one observer to ensure reliability of 0.65 or higher. A recommendation from this study proposed supplementing full-lesson observations with shorter observations by others as a way to save time and control costs. This study also found range restriction in the use of observation instruments, as observers rarely used the top or bottom categories on the four-point scale. Also notable was that compared to peer raters, administrators differentiated more among teachers in their ratings.

A review by Paulsen (2002) had previously noted that, compared to student ratings, the reliability and validity of peer ratings of teaching are not as well established. Research has indicated that peer ratings based solely on classroom observations are not generally reliable (Centra, 1993). There is a general consensus that adequate training and increasing the number of observers and classroom visits in combination increase the reliability of peer observation to acceptable levels (Braskamp & Ory, 1993; Centra, 1993; Paulsen, 2002).

In summary, regardless of how information from observations are used, authors and researchers recommend that *what* is to be observed be carefully defined ahead of time, that observations be made at multiple points in time, and that multiple observers with proper training be used (Goe et al., 2008; Kane et al., 2012; Oliva et al., 2009). Thus, a potential drawback of classroom observations is the cost of conducting them properly, as personnel time, training, and calibration and/or certification are required. Observations are only as good as the instruments used, and the usefulness of an instrument is dependent upon observer training and its proper application. Authors note that in the absence of either sound training or an adequate numbers of observers, peer ratings based solely on classroom observation are not generally reliable (Centra, 1993; Kane et al., 2012; Paulsen, 2002). Further, a review by Goe et al. (2008) noted evidence that training for principal evaluations remains limited and rare, a factor that impairs validity.

Research generally supports the use of observations for formative purposes (Goe, Bell & Little, 2008). There is empirical support that, when conducted with the proper infrastructure (e.g., validated instruments, rater training and certification, multiple observations by multiple observers), classroom observations can be valid and reliable enough for high-stakes decision making (Ho & Kane, 2013; Kane et al., 2012). Notably, research on classroom observation often involves the use of pre-recorded videos of teachers in classroom instructional settings. Aside from research applications, the video method shows great potential for teacher feedback and for the training and assessment of observers (Bill and Melinda Gates Foundation, 2013).

Observations of classroom teaching are currently utilized in instructor certification and evaluation practices in the Army and other Services. The use of senior or master instructors to evaluate other instructors is beneficial, as these evaluators hold the relevant pedagogical and content knowledge required of the role. Army instructional settings are also well-suited for peer evaluation of teaching. Peer instructors are also most likely to hold the required expertise (i.e., domain knowledge and pedagogical knowledge) to conduct meaningful assessments. However, while a checklist approach can standardize *what* is being observed, instruments are only as thorough as what they are designed measure. Checklists that focus on trivial and easy-to-measure aspects of an instructor's practice and other classroom factors likely result in missed opportunities for formative assessment and feedback. A summative score or Go/No Go determination may satisfy an institutional metric, but contextual feedback that has meaning to the observed instructor is important to ensure the results have formative value.

Student achievement and value-added modeling. Measures of student achievement and growth represent learning (Level 2) in Kirkpatrick's four-level training evaluation model (1994). The focus of this type of measurement is on outcomes of the training (or teaching) and not the inputs (i.e., instructor's use of strategies and techniques in the classroom). As an example, administrators may use student achievement scores to examine student performance at the end of a school year, which is useful for determining the percentage of a class, grade or school that meets a given standard (Aaronson, Barrow & Sander, 2007; Hershberg, Simon & Kruger, 2004). Thus, in relation to this effort's definition of an effective instructor, student achievement and growth gains are useful for measuring instructor effectiveness in creating positive student outcomes, but not for the instructor's use of appropriate instructional tools (strategies and techniques) and in demonstrating empathy and a personal capability to tailor instruction based on individual differences.

An increasingly common measure of student achievement is value-added modeling, which examines student growth in learning. This approach involves the application of complex statistical techniques that use multiple years of student test score data to estimate the effects of individual schools or teachers (McCaffrey, Lockwood, Koretz, Louis & Hamilton, 2004). Value-added modeling is also referred to as teacher or school effects, growth measures, or yearly progress or growth. As a method, value-added modeling is useful in providing a summary score of the contribution of various factors toward growth in student achievement (Goldhaber & Anthony, 2004). The robust metric allows administrators to make informed, data-driven decisions and to focus resources to aid student progress. Administrators also use the metrics to benchmark against other schools or districts and to direct teacher attention and focus on student growth as opposed to achievement. Value-added modeling is distinct from other measures of

student achievement in that it examines student growth over a year. Historical student data are used to establish a performance baseline and calculations determine growth relative to students' baseline performance. The results show the contribution of a teacher or school to student growth (Aaronson et al., 2007; Hershberg et al., 2004; McCaffrey et al., 2004).

There is limited research on the validity and reliability of student achievement scores and value-added models. Correlating value-added scores and teacher qualifications, characteristics or practices has yielded mixed results (Goe et al., 2008). Some studies have concluded that the value-added methodology is neither fair enough, nor reliable enough, nor valid enough to be used as a basis for high-stakes decisions about teachers (National Education Association, 2010). Researchers typically report the reliability of value-added measures in the range of 0.30 to 0.50, which is higher than what studies have found for a single classroom observation alone (0.14 to 0.37) (Kane et al., 2012).

While increasingly common, value-added modeling is controversial and the least understood by most education professionals and teachers (Goe et al., 2008). A major drawback with using student achievement measures such as value-added modeling is that it does not provide an understanding of *what* effective teachers do that makes them effective (Rivkin, Hanushek & Kain, 2005). The approach is outcomes-based and does not provide any information about a teacher's performance, specifically the instructional strategies and techniques used by the teacher (teacher processes). Thus, these measures lack formative value. When most students in a class perform better than predicted on standardized achievement tests, the teacher is credited with being an effective teacher (and vice versa). Research suggests teachers differ substantially in their contributions to students' test score gains (Goe et al., 2008). Also problematic is that value-added models assume teachers are the sole influence on student achievement rather than considering other factors that contribute to student outcomes (e.g., schools, family, peers) (McCaffrey et al., 2004). No single teacher accounts for all of a student's learning, and it is impossible to fully identify the influence of additional factors that affect student performance (National Education Association, 2010).

The use of student achievement scores and value-added modeling to measure effective instruction in Army settings is likely constrained for several reasons. First, the *pass rate* for Soldiers attending Army courses, particularly within NCOES, is often very high. Student achievement on written and practical examinations is largely determined by Go/No Go determinations that are based on a threshold (e.g., 70%). High pass rates do not generate sufficiently robust scores for meaningful analysis of student achievement. However, at a holistic level, an unusually high or low pass rate for a class compared to other classes in an academy may be an indication of effective or ineffective instruction worth further diagnosing. Applying a value added methodology would likely require a pre- and post-test measure to determine student growth during the course, as students would not enter a course with baseline scores for comparison. The nature of the examination and test scoring is not likely to generate sufficiently robust data points to make a value-added analysis meaningful. Regardless, student achievement scores do not offer useful information on *how* instructors help students achieve various outcomes. Student achievement is also influenced by factors other than an effective instructor, including any independent learning, peer learning, and interaction with other instructor cadre that occurs during the course.

Student evaluation of teaching. Student evaluations of teaching represent reaction (Level 1) in Kirkpatrick's four-level training evaluation model (1994). Student evaluations often consist of surveys and rating scales completed by learners, used to gather opinions or judgments about teaching practice. Students are a logical source of information, as they have the most direct contact with teachers and are the direct consumers of teaching processes. In relation to this effort's definition of an effective instructor, student evaluation of teaching provides useful information for measuring instructor effectiveness in demonstrating empathy and a personal capability to tailor instruction based on individual differences. Student evaluations are less useful for providing information about an instructor's effectiveness in applying appropriate instructional tools (strategies and techniques) and in creating positive student outcomes (other than student satisfaction).

While a common practice, it has been noted that student ratings are rarely taken seriously as part of teacher evaluation systems (Goe et al., 2008). However, Paulsen (2002) notes that student ratings play a dominant role in the operational definition of what constitutes effective teaching for faculty in institutional settings. In a review of student evaluation of college teaching effectiveness, Wachtel (1998) acknowledges an enormous amount of literature and studies on student evaluations of instruction. Even two decades ago, Marsh and Dunkin (1992) estimated the number of papers to be in the thousands.

Studies have demonstrated the reliability of student ratings to be generally robust (Cashin, 1995; Feldman, 1977; Follman, 1992) though ratings tend to skew favorably (Worrell & Kuterback, 2001). A review by Paulsen (2002) summarized results of meta-analyses that found reliability coefficients for interrater agreement of about .70 or higher when more than ten raters are surveyed on well-established rating forms (Cashin, 1995; Centra, 1993). A meta-analysis by Feldman (1989b) cited positive correlations between student ratings and the ratings of others, including alumni (.69), colleagues (.55), administrators (.39), third-party trained observers (.50), and instructors themselves (.29). Additionally, qualitative evaluations by students have been found to be highly correlated with their quantitative ratings of teaching effectiveness (Braskamp & Ory, 1994).

There are persistent validity concerns with student evaluations of teaching due to potential biases that affect ratings (e.g., leniency, halo) and students' lack of knowledge about the full context of teaching (Follman, 1992; Goe et al., 2008). However, student ratings have been found to positively correlate with measures of student achievement (Kyriakides, 2005; Wilkerson, Manatt, Rogers & Maughan, 2000). Researchers have reported moderate to strong (.30 to .50) correlations between student ratings and student performance on final examinations (Cohen, 1981; Feldman, 1989a).

Numerous studies have concluded that student ratings are valid, reliable and worthwhile as a means for evaluating teaching (Centra, 1993; Cohen, 1981; Marsh & Dunkin, 1992; Paulsen, 2002; Wachtel, 1998). While useful as a component of teacher evaluation, student evaluations should not be a primary or sole criterion. An important implication of the empirical findings is that for summative and/or high-stakes purposes, teacher ratings should be collected from an adequate number of students and should cover different courses and years (Centra, 1993; Marsh

& Dunkin, 1992; Paulsen, 2002).

Instructors often express concerns about the meaningfulness and appropriateness of data from student evaluations (Paulsen, 2002; Turpen, Henderson & Dancy, 2012), often because students are rarely if ever qualified to rate teachers on certain areas of effective teaching, such as content knowledge, curriculum, classroom management, and collegiality (Follman, 1992; Worrell & Kuterbach, 2001). As validity is dependent upon the instrument used and its administration, some experts generally recommend using student evaluations of teaching as formative assessments only (Goe et al., 2008). Feedback from student ratings can offer formative value to improve teaching practice, though the feedback alone will not automatically improve teaching and sustain improvement without other types of feedback (Wachtel, 1998). Timing is important when student ratings are used for formative purposes. Conducting student evaluations earlier in a course allows instructors an opportunity to improve. Research has indicated that the time at which student evaluations are administered does not have an effect on the results (Feldman, 1979).

Student evaluations are a current element of Army instructional practices. A hallmark of Army training and education is the after action review (AAR), a process which seeks input and feedback from participants in the training audience. Similarly, the Army and other Services solicit student feedback as part of after course evaluations. Instructors and cadre are typically rated as a component of these evaluations. Thus, there is opportunity to expand upon current instructor rating practices to include greater levels of detail for which students are an appropriate source of information (e.g., ability to tailor instruction based on individual differences).

Self-assessment. Self-assessments represent a teacher's report of how well he or she is working with students in and outside of the classroom. The most useful self-assessments capture the teacher's beliefs, intentions and expectations and assess strengths and areas for growth. Self-assessments are by nature subject to bias given they are based on self-reported data (Oliva et al., 2009). Self-assessments are useful in that they can provide introspective indicators toward measuring instructor effectiveness across this effort's definition of effective instruction. An instructor can assess his or her own abilities in applying instructional tools, in demonstrating empathy and a personal capability to tailor instructor to meet student needs, and in creating positive student outcomes. However, self-assessment alone is not a sufficient measure for assessing or evaluating instructor effectiveness.

Authors tend to agree that self-evaluations lack the validity and objectivity necessary for summative evaluation (Centra, 1993; Paulsen, 2002), and are insufficient as a standalone measure of effective teaching. Rather, self-assessments provide a useful perspective for comparison to other ratings (e.g., classroom observations, student evaluations) or performance data (e.g., student achievement or growth). Perhaps more appropriately, when compared to other measures of instructor effectiveness, results of a self-assessment may reveal an instructor's *blind spot* for formative improvement. As with other measures for evaluating teaching effectiveness, evaluations systems should utilize carefully designed and validated self-assessment instruments for their intended purpose.

Portfolios. Teacher portfolios are collections of materials (i.e., instructional artifacts) compiled by an instructor to exhibit evidence of teaching practice, course activities and student progress. Examples of instructional artifacts include lesson plans, assessments, curriculum design, student work samples, communications, videos of classroom instruction, and reflective writing (Darling-Hammond & Snyder, 2000). Teaching portfolios are commonly used in teacher preparation, licensure and certification programs, and also as a component of teacher selection practices. Painter (2001) suggests that portfolios include both teacher and student work, selected through thoughtful reflection so as to avoid compiling a teaching ‘scrapbook.’ The (reflective) writing component is also important, as the process often requires an instructor’s defense as to why an artifact is included in the portfolio and how it relates to standards of teaching. In relation to this effort’s definition of an effective instructor, teaching portfolios may provide indicators of effective teaching across all three elements of the definition (i.e., applying appropriate instructional tools, creating positive student outcomes, and tailoring instruction based on individual differences), depending on the artifacts included. However, like self-assessments, a teaching portfolio alone is not a sufficient measure for assessing or evaluating instructor effectiveness.

A study by Tucker, Stronge, Gareis and Beers (2003) found that portfolios were able to document the fulfillment of 18 teacher-performance responsibilities covering four domains (instruction, assessment, management, and professionalism) specified by a school division’s evaluation system. In the study’s sample, 90% of portfolio artifacts demonstrated content validity (i.e., relevance to one or more of the teacher responsibilities). On average, teachers included 24 valid artifacts in their portfolios. About half of a typical portfolio’s artifacts addressed the domain *professionalism* (e.g., committee work, communications with parents) while one-fifth of the artifacts addressed *instructional responsibilities*. The portfolios included relatively fewer artifacts addressing the domains of *assessment* and *classroom management*. However, these findings illustrate the positive role of portfolios in documenting *professionalism* and *assessment*, two aspects of teacher performance not easily observable by administrators during classroom observations or in informal settings.

Goe et al. (2008) note that while portfolios offer a comprehensive and in-depth portrait of teaching practice, the complexity raises concerns about reliability of evaluating them. Studies on the interrater reliability of large-scale portfolio assessments have found the percentage of agreement is usually between 45 percent and 75 percent with correlations between raters rarely reaching 0.80, lower than desirable for high-stakes decision making (Johnson, McDaniel & Willeke, 2000). Paulsen (2002) notes the research on the reliability of peer review of portfolios appears promising. In a small-scale study that involved peer evaluation of faculty dossiers, composite reliability coefficients of six evaluators were .90 and higher across the areas of research, teaching and service (Root, 1987).

The study by Tucker et al. (2003) concluded that portfolios do enhance the evaluation of teachers for both accountability and professional development purposes. Portfolios provide evidence of teacher practice that are less easily measured through other means such as classroom observation, though authors Goe et al. (2008) note there is a lack of research linking portfolios to actual student achievement. Teachers and administrators tend to view portfolios as fair and accurate, though teachers express concerns about feasibility, as a potential drawback of

portfolios is the time required to compile the materials. When used for teacher evaluation, portfolios are meant to exhibit exemplary work (i.e., fulfillment of predetermined standards). Thus, they are subject to bias, as teachers decide what to include (Oliva et al., 2009). As with other measures of effective teaching, authors recommend the use of portfolios inclusively but not exclusively in the evaluation of teachers (Goe et al., 2008; Tucker et al., 2003).

Portfolios are not currently a formal component of instructor professional development or evaluation in the Army or other Services. However, portfolios have potential utility in Army instructional settings. An Army instructor could maintain a portfolio and add to it throughout his/her instructor career or assignments, including instructional artifacts of individual preparation and training, assessments, evaluations, student work, and exemplar activities.

Considerations for Instructor Identification, Preparation and Evaluation

The results of the determining which instructor process, among Identification, Preparation and Evaluation, is appropriate for assessing or developing each instructor KSAO and work behavior are presented in Table 7. In general, those KSAOs and work behaviors that are not easily learned were assigned to Identification, while those KSAOs and work behaviors that can be learned were assigned to Preparation. One specific result of the classification exercise was that Evaluation was not selected as the best process for identifying instructor KSAOs and work behaviors. Rather, it was determined assessment of KSAOs are important for providing formative feedback to instructors, while evaluation of KSAOs provide measures of instructor effectiveness once the instructor was performing on the job. The project team also felt that Evaluation of the KSAOs was important to provide feedback on the job performance of the instructor but the Identification and Preparation processes were critical to ensure the right instructors with the right KSAOs were put into a position to affect student development. Evaluation (e.g., measurement) is also a subcomponent of Identification and Preparation, so it is difficult to distinguish it as an orthogonal process. Therefore, in Table 7, the Evaluation Methods column entries reflect for each row how best to assess or evaluate that KSAO and work behavior during the identification or preparation process, as appropriate.

Table 7.

KSAOs and Work Behaviors Linked to Appropriate Processes.

KSAO	Work Behavior	Process	Identification Methods	Preparation Methods	Evaluation Methods
Knowledge of subject matter being taught	Maintain expertise in topic area	Identify and Prepare	Qualifications, Interview	Reading, Lecture, Discussion, Problem Solving	Inherent in Identify/Prepare phases; Assessment by Supervisor, Peer (Observation, Portfolio)
Knowledge of traits and behaviors of adult learners	Select/implement instructional strategies and techniques	Prepare		Reading, Lecture, Discussion, Problem Solving	Inherent in Identify/Prepare phases; Assessment by Supervisor, Peer (Observation, Portfolio)
Knowledge of student's current level of performance	Evaluate student performance	Prepare		Reading, Lecture, Discussion, Problem Solving	Assessment by Supervisor, Peer (Observation, Portfolio)

KSAO	Work Behavior	Process	Identification Methods	Preparation Methods	Evaluation Methods
Knowledge of principles and methods for curriculum and training design	Select/implement instructional strategies and techniques	Prepare		Reading, Lecture, Discussion, Problem Solving	Inherent in Identify/Prepare phases; Assessment by Supervisor, Peer (Observation, Portfolio)
Knowledge of principles and methods of teaching individuals and groups	Select/implement instructional strategies and techniques	Prepare		Reading, Lecture, Discussion, Problem Solving	Inherent in Identify/Prepare phases; Assessment by Supervisor, Peer (Observation)
Knowledge of principles and methods for assessing for training effectiveness	Evaluate student performance	Prepare		Reading, Lecture, Discussion, Problem Solving	Inherent in Identify/Prepare phases
Knowledge of the structure and content of the English language	Present/facilitate course materials	Identify	Qualifications, Interview, Demonstration, Work samples		Inherent in Identify phase
Knowledge of coaching methods and techniques	Mentor/coach students	Prepare	Qualifications (past evaluations), Interview	Reading, Lecture, Discussion, Problem Solving, Role Play	Inherent in Identify/Prepare phases; Student Evaluation of Teaching; Assessment by Supervisor, Peer (Portfolio)

KSAO	Work Behavior	Process	Identification Methods	Preparation Methods	Evaluation Methods
Skill at observing and monitoring students	Monitor/ observe students	Prepare		Reading, Lecture, Problem Solving, Role play, Simulation	Assessment by Supervisor, Peer (Observation)
Skill at employing questioning techniques (e.g., active, open-ended, leadoff) to assess student understanding and/or facilitate discussion	Question students	Prepare		Reading, Lecture, Problem Solving, Role play, Simulation	Assessment by Supervisor, Peer (Observation); Student Evaluation of Teaching
Skill at utilizing active listening to ensure understanding and build on student ideas	Build rapport with students	Identify and Prepare	Qualifications (past evaluations), Interview	Reading, Lecture, Problem Solving, Role play, Simulation	Assessment by Supervisor, Peer (Observation); Student Evaluation of Teaching
Skill at formal and informal assessment to measure student progress on core course content	Evaluate student performance	Prepare		Reading, Lecture, Problem Solving, Role play, Simulation	Assessment by Supervisor, Peer (Observation, Portfolio); Student Achievement Scores/Gains
Skill at making use of multiple instructional strategies and techniques (e.g., scaffolding; blended learning) to account for individual differences in learner behavior/thought processes	Select/implement instructional strategies and techniques	Prepare		Reading, Lecture, Problem Solving, Role play, Simulation	Assessment by Supervisor, Peer (Observation); Student Evaluation of Teaching

KSAO	Work Behavior	Process	Identification Methods	Preparation Methods	Evaluation Methods
Skill at providing formal and informal feedback so students understand strengths and weaknesses	Provide formal and informal feedback	Prepare		Reading, Lecture, Problem Solving, Role play, Simulation	Assessment by Supervisor, Peer (Observation); Student Evaluation of Teaching
Skill at presenting and facilitating course materials to show content in progression and bring students to end goal	Present/facilitate course materials	Identify and Prepare	Qualifications, Demonstration	Reading, Lecture, Problem Solving, Role play, Simulation	Assessment by Supervisor, Peer (Observation, Portfolio); Student Evaluation of Teaching
Skill at mentoring and coaching to develop student leadership skills and motivation	Mentor/coach students	Identify and Prepare	Qualifications (past evaluations), Interview	Reading, Lecture, Problem Solving, Role play, Simulation	Student Evaluation of Teaching; Assessment by Supervisor, Peer (Portfolio)
Skill at applying educational technology in ways that enhance student learning	Select/implement instructional strategies and techniques	Prepare		Reading, Lecture, Problem Solving, Role play, Simulation	Assessment by Supervisor, Peer (Observation, Portfolio); Student Evaluation of Teaching; Student Achievement Scores/Gains

KSAO	Work Behavior	Process	Identification Methods	Preparation Methods	Evaluation Methods
Ability to communicate information and ideas in speaking so others will understand	Present/facilitate course materials	Identify and Prepare	Qualifications, Interview, Demonstration	Lecture, Discussion, Role Play, Presentations/Simulation	Assessment by Supervisor, Peer (Observation); Student Evaluation of Teaching
Ability to communicate information and ideas in writing so others will understand	Present/facilitate course materials	Identify and Prepare	Qualifications, Interview, Demonstration (Writing sample or exercise)	Problem Solving, Reports	Assessment by Supervisor, Peer (Portfolio); Student Evaluation of Teaching
Ability to accurately and effectively interpret students' comments in both verbal and written form	Build rapport with students	Identify and Prepare	Qualifications, Interview, Selection assessment	Role Play, Problem Solving/Exercises	Student Evaluation of Teaching; Assessment by Supervisor, Peer (Portfolio)
Ability to apply learning theory to individual instructional circumstances	Plan/prepare lessons and activities	Prepare		Lecture, Discussion, Problem Solving/Exercises	Assessment by Supervisor, Peer (Observation)
Ability to combine pieces of information to form general rules or conclusions	Maintain expertise in topic area	Identify	Qualifications, Interview, Selection assessment	Lecture, Discussion, Problem Solving/Exercises	Assessment by Supervisor, Peer (Observation)

KSAO	Work Behavior	Process	Identification Methods	Preparation Methods	Evaluation Methods
Ability to apply general rules to specific problems to produce answers that make sense	Maintain expertise in topic area	Identify	Qualifications, Interview, Selection assessment		Assessment by Supervisor, Peer (Observation)
Openness to experience	All	Identify	Qualifications, Interview, Selection assessment (personality)		Inherent in Identify phase
Low need for control/ Tolerance for ambiguity to allow for classroom discussion and group problem-solving when applicable	Manage student discipline	Identify	Interview, Selection assessment (personality)		Inherent in Identify/Prepare phases; Assessment by Supervisor, Peer (Observation); Student Evaluation of Teaching
Believe students are responsible for and capable of own learning	Mentor/coach students	Identify	Interview		Inherent in Identify/Prepare phases; Student Evaluation of Teaching
Value independent thought	Mentor/coach students	Identify	Interview		Inherent in Identify phase; Student Evaluation of Teaching

KSAO	Work Behavior	Process	Identification Methods	Preparation Methods	Evaluation Methods
View learning as development of independent thinking skills	Question students	Identify	Interview		Inherent in Identify phase
Believe learning is a collaborative process	Mentor/coach students; Select/implement instructional strategies and techniques	Identify (prepare)	Interview	Lecture, Discussion, Problem Solving/Exercises	Inherent in Identify/Prepare phases; Student Evaluation of Teaching
Accept student-centered methods as valid	All	Identify (prepare)	Interview	Lecture, Discussion, Problem Solving/Exercises	Inherent in Identify/Prepare phases
View teaching as a learning profession	Maintain expertise in topic area	Identify (prepare)	Interview	Lecture, Discussion, Problem Solving/Exercises	Inherent in Identify/Prepare phases; Assessment by Supervisor, Peer (Portfolio)

For some KSAOs, both Identify and Prepare processes were selected as equally appropriate. For example, for the KSAO “Skill at mentoring and coaching to develop student leadership skills and motivation,” it was determined that instructors would likely already have exhibited coaching, and to a lesser extent, mentoring behaviors as a result of being in previous unit training or leadership positions. However, it was also felt that additional training and development in coaching and mentoring would be very valuable to ensure best practices in these behaviors are exhibited in Army institutional training assignments.

Both identification and preparation methods proposed are dependent on several factors, including resources available to develop selection instruments and development content, the time available for either selecting or developing instructors, and the effectiveness of the selection or developmental method. The methods proposed for the identification process consisted of:

1. Qualifications – as established through various instruments including OER/NCOERs, resumes, portfolios, and other documentation.
2. Interviews – principally structured interviews.
3. Demonstrations and work samples – structured simulations that would require the candidate to demonstrate the KSAOs under assessment. These would include writing samples and sample lectures.
4. Tests – such as personality and other written tests.

The methods proposed for the preparation process included:

1. Reading – generally self-study of written materials on the KSAO topics or subjects.
2. Lecture – generally learning from others with in-depth knowledge and experience in the topic or subject (KSAO).
3. Discussion – either with a knowledgeable other or within a learning context such as a discussion group or small group.
4. Problem Solving – general term for methods that include a wide range of exercises to engage students in thinking about real world applications of the KSAO and in particular, working through various challenges or problems with instructing.
5. Role Play – methods that encourage the student to demonstrate behaviors during simulated interactions with others, often with scripts or structure to ensure certain behaviors are exhibited.
6. Simulation – other simulations which may include computer based exercises, virtual environments, and games which encourage the students to try out behaviors with the KSAO domain.
7. Reports and other demonstrations – assignments given during instructor preparation that would demonstrate proficiency in the KSAO under assessment.

The methods proposed for instructor evaluation are also dependent on several factors, including the intended purpose of the measurement (i.e., formative or summative), the resources available to develop sound measures (e.g., instruments, protocols, rubrics), and the resources available to conduct evaluations (e.g., personnel time and availability). The methods proposed for instructor evaluation processes included:

1. Classroom observations – conducted by supervisors, instructor peers, or third party observers with relevant instructional and content knowledge and in a position to provide valid judgments of student/instructor performance.
2. Student achievement and value-added modeling – student achievement scores or measures of student growth/gains as assessed by pre- and post- measures, or statistical value-added modeling, representing the results level in training evaluation (Level 2).
3. Student evaluation of teaching – ratings and written evaluations by students of the instructor, representing the reaction level in training evaluation (Level 1).
4. Self-assessment – an instructor’s own assessment of teaching practice, strengths and areas to improve, in the context of teaching requirements for that position.
5. Portfolios – a collection of instructional artifacts (self-prepared) that may include student work products, lesson plans, supporting exercises and exhibits, self- and other assessments, reports and other documentation, and personal reflections of the same.

DISCUSSION

Discussion of the relative merits of approaches to identification, preparation, and assessment of effective instructors begs the definition of “effective instructor.” As an output of the Foundational Task, in parallel with the listing of instructor KSAOs, an operational definition of an effective instructor was developed. This definition, detailed below, serves to frame this discussion of achieving the ALM’s end state for Army instructors.

Operational Definition of an Effective Instructor

Based on the literature review, workshop, and additional input from SMEs, we developed the following operational definition for an effective instructor:

“An effective instructor is one who can, by perceiving the individual differences in students and learning environments and applying instructional strategies and techniques as appropriate for the situation, create positive student outcomes related to the short and long term objectives of a course.”

This definition of instructor effectiveness comprises three interdependent aspects. These include:

- applying appropriate instructional tools (strategies and techniques),
- creating positive student outcomes, and
- demonstrating empathy and a personal capability to tailor instruction based on individual differences.

The definition is more outcome than process based. The outcome is at the level of the individual student and is dependent on how well defined and assessable the course goals and objectives may be. The effective instructor achieves these goals and objectives by tailoring instructional technique to individual students within the constraints of the course.

Some examples of what is meant by “applying instructional strategies and techniques” and “creat[ing] positive student outcomes” can be found in Table 8. Please note that these lists are not comprehensive.

Table 8

Examples of Instructional Strategies and Techniques and Positive Student Outcomes

Instructional strategies and techniques	Positive student outcomes
Observing and monitoring	Knowledge retention
Questioning techniques	Knowledge transfer
Active listening	Student scores from rigorously validated tests
Formal and informal assessment	Student motivation
Formal and informal feedback	Student self-efficacy
Mentoring and coaching	Skill development
Presenting and facilitation	

Effective methods for selecting instructors

Two conclusions can be drawn from consideration of selection practices. The first is the importance of considering multiple sources of information when making the selection decision. According to the selection literature, research in this area has shown that measuring across a job candidate’s skills and abilities is often a better predictor of job performance than focusing on any one measure. In a meta-analysis conducted by Schmitt and Hunter (1998), a general mental ability test in conjunction with either a structured interview (i.e., mean validity of .63) or work sample (i.e., mean validity of .65) was one of the most effective means of predicting a job candidate’s future job performance. In the meta-analysis, job performance was defined as dollar value of output or output as a percentage of mean output, but the finding appears to hold true in teacher selection. Gimbert and Chesley (2009) indicate that rarely does a single measure explain more than 25% of the variance in a teacher’s later job performance.

The second is the fidelity of the measure being used to select the teacher. Fidelity is defined as “the extent to which a predictive measure is similar... to the behavior that is to be predicted” (Schalock, 1979, p.369). The selection literature indicates that the closer the selection measure matches what candidates will actually do on the job (and the environmental conditions of the position), the better it will be at predicting future performance (Webster, 1988; Winter, 1995; Wise et al., 1987). An example of a high fidelity selection measure would be a work sample where the candidate is asked to perform similar tasks to those in an actual classroom (e.g., teaching demonstration). On the other hand, a low fidelity selection measure would be something like a curriculum vita (CV) or letters of reference.

Table 9 shows the level of empirical support for each method discussed above with regard to its ability to identifying effective instructors.

Table 9

Empirical Support for Identification/Selection Method

Method of Identification/Selection	Level of Empirical Support
Intelligence	Low
Level of Education	Low
Personality Traits and Characteristics	Low
Interviews	Medium
Work Samples	High

Effective Methods for Preparing Instructors

The majority of the instructional effectiveness research deals with a single environment, typically a classroom or in some cases distributed environments. Much of the research occurred before the saturation of mobile devices, social media and application of games to learning. Generally speaking, the research supports instructor development methods that emphasize observing and modeling other effective instructors. A number of the evaluations of teacher professional development also focus on the inclusion of colleagues in offering critical reviews of in-class performance and ongoing coaching and mentoring (Meirink et al., 2008).

Instructor preparation methods that have been linked to higher student learning and motivation include problem-based instruction, coaching and mentoring, the use of video and audio replay in critiquing practice, improving the application of knowledge and skills in operational conditions (e.g., experiential and situated learning), and the use of certifications to ensure competence. Other methods which have potential value for preparing Army instructors include using collaboration and cooperative learning to build professionalism, utilizing team training techniques where teams of instructors are used, and leveraging distance or distributed training capabilities to better prepare instructors prior to institutional training or in circumstances that prevent face to face interactions.

Beyond the specific methods and techniques of preparing instructors this section has also touched upon instructional contexts, such as classroom, online and social settings as well as instructional media, including text-based reference materials, video, audio, games and computer simulation. Media and contexts interact with methods and should also be considered when determining optimal opportunities for preparing Army instructors. Obviously resources, including time and cost must also be considered.

Effective Methods for Evaluating Instructors

There are several conclusions that may be drawn from the research on methods for assessing and evaluating instructors. This discussion begins by offering three considerations for evaluation system design.

- First, it is important to determine the purpose of instructor assessment or evaluation

before selecting measures. Whether a measure is formative or summative is dependent upon how the information is used. To have formative value to an instructor, feedback should provide information on how to improve or change teaching behavior. Ideally, evaluation systems should drive effective instruction, not just measure it. Goe et al. (2008) note that it is important to design evaluation systems that use multiple indicators of effective teaching, that differentiate among teachers by what is being taught, and that measure what is important to the institution. It is also important to give teachers opportunities to improve as well as resources and training. Thus, defining foundational concepts such as the KSAOs and work behaviors (WB) that constitute effective instruction are of paramount importance.

- Second, evaluation systems should use carefully designed and validated instruments, and use them for their intended purpose. The validity of a measure or practice is only as good as the instruments used. A simple checklist approach, while potentially valid, tends to offer low quality information about teaching practice. A robust measure (i.e., instrument or rubric) designed to capture specific elements of instructional practice or behaviors offers more value to an institution and to instructors. Observers should be trained on the instruments used, rater reliability should be established, and periodic recalibration should occur (Goe et al., 2008).
- Third, instructor evaluation systems should include multiple measures, used at multiple points in time. No single measure of instructor effectiveness sufficiently captures all of the important elements of effective instruction, and the fewer the indicators, the greater the potential for error (Goe et al., 2008). Increasing the frequency of use of formative measures allows for the results to be used for ongoing professional development opportunities (e.g., goal setting) for teachers (Oliva et al., 2009). Put broadly, these considerations point to the need for proper resourcing when designing instructor evaluation systems. This includes the development of validated instruments, training and certification for evaluators, and the time for multiple evaluators to spend engaged in classroom observations and reviewing teaching artifacts.

Both empirical evidence and expert opinion support the use of several methods for assessing and evaluating instructors. Observations (classroom or video-based) can serve as valid and reliable measures of instructor effectiveness if done properly (e.g., carefully designed observation instrument, use of multiple observers who are trained and calibrated, multiple observations over time). Teaching observations should be conducted by observers who are experienced, have taught the same or similar course, and are trained in peer review process (Ammons & Lane, 2012). Observations are most useful when formative feedback is provided to the observed instructor along with a primer for self-reflection and/or response.

When selecting evaluation methods, it is important to consider the source of the information. Research suggests multiple sources and types of data should be used, and the most common sources are students, peers and teachers themselves. Peer reviewers are especially useful for evaluating an instructor's subject matter mastery and discipline-specific aspects of instructional design and pedagogy (Paulsen, 2002). Students offer a unique vantage point for measuring effective instruction, as students have the most regular and direct contact with the

instructor. Student evaluations of teaching should be a component of teacher evaluations, but not the primary or sole criterion. Students are rarely qualified to rate teachers on areas such as curriculum, classroom management, content knowledge, and collegiality (Follman, 1992; Goe et al., 2008).

Self-assessments or other self-report practices (e.g., portfolio of instructional artifacts) are useful methods for capturing an instructor's perspective on his/her teaching practice. Portfolio assessments should be used inclusively to complement data collected through classroom observation and other sources, not as a stand-alone assessment for decision making processes (Tucker et al., 2003; Johnson et al., 2000). A measure of student achievement or growth is also useful as a measure of positive student outcomes. When possible, growth or gain is preferred over student achievement scores, to measure change in learning rather than learning achievement.

Each of the evaluation methods examined in this review provided evidence of effective instruction as defined by this effort, but no one method sufficiently captures the full picture. Table 10 displays linkages between the three interdependent aspects of an effective instructor and appropriate methods for instructor assessment and evaluation. In some cases, evidence in this review supports the use of a method as merely an *indicator* of effective teaching practice and not a sufficient standalone source. In other cases, the method may be suitable to *assess/evaluate* the practice. For example, student achievement scores and value-added modeling can provide sufficient information for determining whether an instructor is effective in creating positive student outcomes. An instructor's portfolio may include indications of positive student outcomes, but the measure of student achievement or gains are the optimal source.

Table 10

Assessment/evaluation methods' applicability to facets of instructor effectiveness

Method	An effective instructor is one who can...		
	Apply appropriate instructional tools (strategies and techniques)	Create positive student outcomes	Demonstrate empathy and a personal capacity to tailor instruction based on individual differences
1. Classroom observations	Assess/Evaluate		Indicator
2. Student Achievement and Value-added modeling		Assess/Evaluate	
5. Student evaluation of teaching			Assess/Evaluate
3. Self-assessment	Indicator	Indicator	Indicator
4. Portfolios	Indicator	Indicator	Indicator

As discussed in this review, the Army and other Uniformed Services currently evaluate instructors by utilizing established competencies and outcomes, models for effective training evaluation (i.e., Kirkpatrick, 1994), and instruments such as observation checklists, rubrics and other tools. Current instructor evaluation measures used in the Army align with the instructor competencies outlined in TRADOC Regulation 600-21. While evaluating the effectiveness of the Army's current instructor evaluation system was not a component of this research, a few observations are made. First, the Instructor Development and Recognition Program (IDRP) is described as voluntary, and it is unclear what methods of evaluation are regularly occurring or required for instructors. Second, evaluation measures using a checklist approach likely do not offer robust information that is useful to the instructor for formative improvement. Third, the use of instructor competencies can be limiting. The foundational task of this research aimed to advance the understanding of instructor effectiveness beyond the competency level to include specific KSAOs and WBs that reflect effective instruction. Thus, a conclusion of this research is that these elements of effective teaching may be used to create robust measures (e.g., instruments, protocols, rubrics) for assessing and evaluating instructors.

Considerations for Instructor Identification, Preparation and Evaluation

The proposed framework was developed to further inform existing Army instructor selection, training and evaluation processes by providing the dimensions (e.g., KSAO and work behavior) as well as specific methods and opportunities for identification, development and

evaluation of instructors across an array of institutional programs. Individual Army training programs can use the framework to determine the extent with which current selection, training and evaluation processes are aligned with best practices and potentially develop additional methods and techniques for greater coverage of requisite KSAO measurement. The framework also supports ongoing efforts to improve learner-centric instructor skills and provides a basis for evaluating instructors on specific KSAOs and work behaviors that are supportive of learner-centric skills.

Overall

This project has sought to provide additional research-based guidance on optimal Army instructor KSAOs and work behaviors irrespective of training or educational course content and contexts. The foundational task which identified the critical instructor KSAOs and work behaviors provides a common set of instructor dimensions that are directly relevant to instructor quality and effectiveness and success on the job across a broad range of courses and instructional contexts.

The KSAOs and work behaviors are more an initial baseline than an exhaustive set. During review of this document, additional KSAOs more specific to the Army institutional education environment were proposed, such as

- Collaborate with fellow cadre to sustain instructional excellence under changing conditions
- Understand and act on curricular intent if lesson design/development falls short of goal
- Avoid biased judgment in evaluation of outside perspectives

The implication of these observations is that as the role of Army instructors evolves, the supporting KSAOs must also change.

The value of having a framework of identification, preparation and evaluation methods is to better inform the development of specific selection methods and instruments, preparation approaches, techniques and course materials, and evaluation instruments and practices.

The results expand upon the current information on instructor competencies, recognition, training and education, and assessment instruments provided in TR 600-21 in several ways. First, this report provides the empirical support for proposed selection, preparation and evaluation methods to allow users to better understand method development and effectiveness. Second, this report describes alternative instructor effectiveness elements, namely KSAOs and work behaviors rather than competencies. These more micro level elements may have added value for instructor identification, preparation and evaluation in terms of helping instructors and instructor systems focus on more discrete behavioral elements. Third, this effort provides information on specific methods for selecting, preparing and evaluating instructors that provide greater detail than existing Army instructor doctrine.

As Army instructors learn new skills and behaviors consistent with learner centered techniques, it is important that they also practice the methods they will be using as facilitators of knowledge and skill transfer. For example, existing NCO training courses may have the content related to teaching instructors about social learning or avatar-based feedback techniques, but they may not actually use social learning or avatar-based techniques. To best understand the learning environment, resources, techniques and student characteristics, Army instructors should immerse themselves in the contexts and methods that they will be expected to employ in the coming years.

References

- Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, 25, 95-135.
- Air Education and Training Command (2012). *Faculty development and master instructor programs*. (AETC Instruction 36-2202). Retrieved from: www.e-publishing.af.mil
- Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology*, 103(1), 1-18.
- Ammons, J. L., & Lane, S. J. (2012). Making teaching visible: Sharing and evaluating using peer observation. *Allied Academies International Conference: Proceedings of the Academy of Educational Leadership*, 17(1), 77-81.
- Arthur, W., Bennett, W., Edens, P. S., & Bell, S. T. (2003). Effectiveness of training in organizations: Analysis of design and evaluation features. *Journal of Applied Psychology*, 88(2), 234-245.
- Austin, R., Smyth, J., Rickard, A., Quirk-Bolt, N. & Metcalfe, N. (2010). Collaborative digital learning in schools: Teacher perceptions of purpose and effectiveness. *Technology, Pedagogy and Education*, 19, 327–343.
- Bai, H. (2012). Students' use of self-regulator tool and critical inquiry in online discussions. *Journal of Interactive Learning Research*. 23(4), 209-225.
- Bangs, J. (2011). Experiential learning in an organizational leadership program. *Journal of College Teaching & Learning*, 8(10), 29–33.
- Beal, S. A., Wright, K., & Topaz, D. (2009). *The use of a multiplayer game to execute light infantry company missions*. (Research Report 1915). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Bell, M. (2002). *Peer observation of teaching in Australia*. Retrieved from http://learningandteaching.vu.edu.au/teaching_practice/improve_my_teaching/evaluation_support_for_my_teaching/Resources/id28_Peer_Observation_of_Teaching_in_Australia.pdf
- Benson, S. (2012). The Relative Merits of PBL (Problem-Based Learning) in University Education. *US-China Education Review*, 4, 424-430. Retrieved from <http://0www.eric.ed.gov.libcat.uafs.edu/PDFS/ED533570.pdf>
- Berry, L., Collins, G., Copeman, P., Harper, R., Li, L., & Prentice, S. (2012). Individual consultations: Towards a 360-degree evaluation process. *Journal of Academic Language & Learning*, 6(3), A16-A35. Spacing issue
- Betts, J. R., Zau, A. C., & Rice, L. A. (2003). *Determinants of student achievement: New evidence from San Diego*. San Francisco: Public Policy Institute of California.

- Bill and Melinda Gates Foundation. (2013). *Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET project's three-year study*. (Policy and Practice Brief), Measures of Effective Teaching Project. Seattle, WA: Author.
- Bigelow, R. M. (2002). *Preservice mentoring: Voices of mentors and proteges*. Unpublished Ph.D. dissertation University of Wyoming, Laramie, WY.
- Blank, R.K., de las Alas, N., & Smith, C. (2008, February). *Does teacher professional development have effects on teaching and learning?* (Report prepared for the Council of Chief State School Officers from the National Science Foundation, Grant # REC 0438358) Retrieved from <http://hub.mspnet.org/index.cfm/15474>
- Bonk, C. J. & King, K. S. (Eds.), (1998). *Electronic collaborators: Learner-centered technologies for literacy, apprenticeship, and discourse*. Mahwah, NJ: Lawrence Erlbaum.
- Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher preparation and student achievement. *Education Evaluation and Policy Analysis*, 31(4), 416-440.
- Boyle, T., Bradley, C., Chalk, P., Jones, R., & Pickard P. (2003). Using blended learning to improve student success rates in learning to program. *Journal of Educational Media*, 28 (2-3), 165-178.
- Brandt, C., Mathers, C., Oliva, M., Brown-Sims, M., & Hess, J. (2007). Teacher evaluation policies in the Midwest Region: Examining district guidance to schools (Issues & Answers Report, REL 2007-No. 030). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Midwest. Retrieved from http://ies.ed.gov/ncee/edlabs/regions/midwest/pdf/REL_2007030.pdf
- Braskamp, L. A., & Ory, J. C. (1994). *Assessing faculty work: Enhancing individual and institutional performance*. San Francisco: Jossey-Bass.
- Brown, A., Bransford, R. Ferrara., R., & Campione, J. (1983). Learning, Remembering, and Understanding. In P. Mussen (Ed.), *Handbook of child psychology: Cognitive development* (pp. 77-166). New York, John Wiley and Sons.
- Callinan, M., & Robertson, I. T. (2000). Work sample testing. *International Journal of Selection and Assessment*, 8(4), 248-260.
- Canale, A. M., Herdklotz, C., & Wild, L. (2012). *Evaluation of teaching effectiveness: Benchmark report & recommendations*. Retrieved from: http://www.rit.edu/academicaffairs/facultydevelopment/sites/rit.edu.academicaffairs.facultydevelopment/files/docs/Evaluation_of_Teaching_Effectiveness.pdf
- Cannon-Bowers, J. & Salas, E. (2000). *Making decisions under stress: Implications for individual and team training*. Washington, DC: American Psychological Association:

- Cantrell, S., Fullerton, J., Kane, T. J., & Staiger, D. O. (2008). *National board certification and teacher effectiveness: Evidence from a random assignment experiment* (NBER Working Paper 14608). Cambridge, MA: National Bureau of Economic Research.
- Carman, J. M. (2005). *Blended learning: Five key ingredients*. Retrieved from:
<http://www.agilantlearning.com/pdf/Blended%20Learning%20Design.pdf>
- Carpenter, T. D., & Wisecarver, M. W. (2004). *Identifying and validating a model of interpersonal performance dimensions*. (Technical Report 1144). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Cashin, W. E. (1995). *Student ratings of effective teaching: The research revisited*. (Idea Paper no. 32). Manhattan, KS: Center for Faculty Evaluation and Faculty Development.
- Cashin, W. E. (1996). *Developing an effective faculty evaluation system*. (Idea Paper). Manhattan, KS: Center for Faculty Evaluation and Faculty Development.
- Centra, J. A. (1993). *Reflective faculty evaluation: Enhancing teaching and determining faculty effectiveness*. San Francisco: Jossey-Bass.
- Cheng, M. I., & Dainty, R. I. J. (2005). Toward a multidimensional competency-based managerial performance framework: A hybrid approach. *Journal of Managerial Psychology*, 20, 380–396.
- Cianciolo, A. T., Grover, J., Bickley, W. R., & Manning, D. (2011). *Problem-based learning: Instructor characteristics, competencies, and professional development*. (Research Report 1936). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). *How and why do teacher credentials matter for student achievement?* (NBER Working Paper 12828). Cambridge, MA: National Bureau of Economic Research.
- Cohen, D. K., & Hill, H. C. (1998). *Instructional policy and classroom performance: The mathematics reform in California* (CPRE Research Report Series RR-39). Philadelphia: Consortium for Policy Research in Education. Retrieved from
http://www.cpre.org/images/stories/cpre_pdfs/rr39.pdf
- Cohen, P. A., (1981). Student ratings of instruction and student achievement: A meta-analysis of multisector validity studies. *Review of Educational Research*, 51(3), 281-309.
- Constantine, J., Player, D., Silva, T., Hallgren, K., Grider, M., Deke, J., & Warner, E. (2009). *An evaluation of teachers trained through different routes to certification: Final report*. (NCEE 2009-2043). Washington, DC: National Center for Education Evaluation and Regional Assistance.
- Conway, M., & Cassidy, M. F. (2001, March). *Evaluating trainer effectiveness*. Info-line, Alexandria, VA: American Society for Training & Development.

- Cooper, W., Leibrecht, B. C., & Lickteig, C. W. (2010). *Instructor's peer-to-peer learning guide for the Army reconnaissance course*. (Research Product 2011-02). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- D'Augostino, J. V., & Powers, S. J. (2009). Predicting teacher performance with test scores and grade point average: A meta-analysis. *American Education Research Journal*, 46(1), 146-182.
- Damon, W. (1984). Peer education: The untapped potential. *Journal of Applied Developmental Psychology*, 5, 331–343.
- Dana, J., Dawes, R., & Peterson, N. (2013). Belief in the unstructured interview: The persistence of an illusion. *Judgment and Decision Making*, 8(5), 512-520.
- Darling-Hammond, L. (1999). *Teacher quality and student achievement: A review of state policy evidence*. Seattle, WA: University of Washington Center for the Study of Teaching and Policy.
- Darling-Hammond, L., & Snyder, J. (2000). Authentic assessment of teaching in context. *Teaching and Teacher Education*, 16, 523-545.
- Darling-Hammond, L., Holtzman, D. J., Gatlin, S. J., & Vasquez Heilig, J. (2005). Does teacher preparation matter? Evidence about teacher certification, Teach for America, and teacher effectiveness. *Education Policy Analysis Archives*, 13(42). Retrieved from <http://epaa.asu.edu/epaa/v13n42/>
- Darling-Hammond, L., Newton, S. P., & Wei, R. C. (2013). Developing and assessing beginning teacher effectiveness: The potential of performance assessments. *Education, Assessment, Evaluation, and Accountability*, 25, 179-204.
- Darling-Hammond, L., Wise, A. E., & Pease, S. R. (1983). Teacher evaluation in the organizational context: A review of the literature. *Review of Educational Research*, 53(3), 285-328.
- Dawn, S., Dominguez, K. D., Troutman, W.G., Bond, R., & Cone, C. (2011). Instructional scaffolding to improve student's skills in evaluating clinical literature. *American Journal of Pharmaceutical Education*, 75(4), 1-8.
- Dick, W., & Carey, L. (1996). *The Systematic Design of Instruction* (4th ed.). New York: Harper Collins College Publishers.
- Dochy, F., Segers, M., Bossche, P. V., & Gijbels, D. (2003). Effects of problem-based learning: a meta-analysis. *Learning and Instruction*, 13(5), 533-568.
- Dowling, C., Godfrey, J. M., & Gyles, N. (2003). Do hybrid flexible delivery teaching methods improve accounting students' learning outcomes? *Accounting Education*, 12, 373–391.

- Duckworth, A. L., Quinn, P. D., & Seligman, M. E. P. (2009). Positive predictors of teacher effectiveness. *The Journal of Positive Psychology*, 4(6), 540-547.
- Dunst, C. J., Trivette, C. M., & Hamby, D.W. (2010). Meta-analysis of the effectiveness of four adult learning methods and strategies. *International Journal of Continuing Education and Lifelong Learning*. 3(1), 92-112.
- Eurich, T. L., Krause, D. E., Cigularov, K., & Thornton, G. C. (2009). Assessment centers: Current practices in the United States. *Journal of Business Psychology*, 24, 387-407.
- Feldman, K. A. (1977). Consistency and variability among college students in rating their teachers and courses: A review and analysis. *Research in Higher Education*, 6(3), 223-274.
- Feldman, K. A. (1979). The significance of circumstances for college students' ratings of their teachers and courses. *Research in Higher Education*, 10, 149-172.
- Feldman, K. A. (1989a). The association between student ratings of specific instructional dimensions and student achievement. *Research in Higher Education*, 30(6), 583-645.
- Feldman, K. A. (1989b). Instructional effectiveness of college teachers as judged by teachers themselves, current and former students, colleagues, administrators, and external (neutral) observers. *Research in Higher Education*, 30(2), 113-135.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34, 906 - 911.
- Follman, J. (1992). Secondary school students' ratings of teacher effectiveness. *The High School Journal*, 75(3), 168-178.
- Gallagher, H. A. (2004). Vaughn Elementary's innovative teacher evaluation system: Are teacher evaluation scores related to growth in student achievement? *Peabody Journal of Education*, 79(4), 79-107.
- Galvao, J. R., Martins, P. G., & Gomes, M. R. (2000). Modeling reality with simulation games for a cooperative learning. *Proceedings of the IEEE 2000 Winter Simulation Conference*, 1692-1698.
- Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical thinking, cognitive presence, and computer conferencing in distance education. *American Journal of Distance Education*, 15(1), 7-23.
- Gimbert, B. G., & Chelsey, D. (2009). Predicting teacher success using teacher selection practices and classroom performance assessment. *Journal of School Leadership*, 19(1), 49-80.

- Glazerman, S., & Seifullah, A. (2012). *An evaluation of the Chicago Teacher Advancement Program (Chicago TAP) after four years*. (Report prepared for The Joyce Foundation). Washington, DC: Mathematica Policy Research.
- Glover, R. W., & Bilginsoy, C. (2005). Registered apprenticeship training in the US construction industry. *Education + Training*, 47, 337-349.
- Goe, L. (2007). *The link between teacher quality and student outcomes*. Washington, DC: National Comprehensive Center for Teacher Quality.
- Goe, L. G., Belle, C., & Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis*. Washington, DC: National Comprehensive Center for Teach Quality.
- Goldhaber, D., & Anthony, E. (2004). *Can teacher quality be effectively assessed?* Washington, DC: Urban Institute. Retrieved from http://www.urban.org/UploadedPDF/410958_NBPTSOutcomes.pdf
- Goldhaber, D., & Brewer, D. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis*, 22(2), 129-145.
- Guarino, C. M., Santibanez, L., & Daley, G. A. (2006). Teacher recruitment and retention: A review of the recent empirical literature. *Review of Educational Research*, 76, 173-208.
- Hague, S. & Srinivasan, S. (2006). A meta-analysis of the training effectiveness of virtual reality surgical simulator. *IEEE Transactions on Information Technology in Biomedicine*, 10(1), 51-58.
- Haney, A. (1997). The role of mentorship in the workplace. In M. C. Taylor (Ed.), *Workplace education* (pp. 211-228). Toronto, Ontario: Culture Concepts.
- Harlen, W. & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education*, 4(3), 365-379.
- Harris, D. N., & Sass, T. R. (2007). *Teacher training, teacher quality and student achievement*. (Working Paper #3). National Center for Analysis Longitudinal Data in Education Research. Retrieved from www.caldercenter.org/PDF/1001059_Teacher_Training.pdf
- Hays, R. T., Jacobs, J. W., Prince, C., & Salas, E. (1992). Requirements for future research in flight simulation training: Guidance based on a meta-analytic review. *The International Journal of Aviation Psychology*, 2, 143-158.
- Heinz, M. (2013). Tomorrow's teachers – selecting the best: An exploration of the quality rationale behind academic and experiential selection criteria for initial teacher education programmes. *Education, Assessment, Evaluation, and Accountability*, 25, 93-114.
- Hershberg, T., Simon, V. A., & Kruger, B. L. (2004). The revelations of value-added. *The School Administrator*, 61, 10-14.

- Hilbert, J., Preskill, H., & Russ-Eft, D. (1997). Evaluating training. In L. J. Bassi & D. Russ-Eft (Eds.), *What works: Assessment, development, and measurement* (pp. 109-150). Alexandria, VA: American Society for Training and Development.
- Hmelo-Silver, C. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review*, 16, 235-266.
- Ho, A. D., & Kane, T. J. (2013). *Reliability of classroom observations by school personnel*. (Research report). Measures of Effective Teaching Project. Seattle, WA: Bill & Melinda Gates Foundation.
- Hobson, L.D., Harris, D., Buckner-Manly, K., & Smith, P. (2012). The importance of mentoring novice and pre-service teachers: Findings from a HBCU student teaching program. *Educational Foundations*, 26(3-4), 67-80.
- Hoole, E. R., & Martineau, J. W. (2014). Evaluation methods. In D. V. Day (Ed.), *The Oxford Handbook of Leadership and Organizations* (pp. 167-196), New York: Oxford University Press.
- Johnson, D. W., & Johnson R. (1999). *Learning together and alone: Cooperative competitive, and individualistic learning*, (5th ed). Allyn & Bacon: Boston.
- Johnson, D. W., & Johnson, R. T. (1988). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher*, 38(5), 365 – 379.
- Johnson, D., & Johnson, R. (1994). *Learning together and alone, cooperative, competitive, and individualistic learning*. Needham Heights, MA: Prentice Hall.
- Johnson, R. L., McDaniel, F. II, & Willeke, M. J. (2000). Using portfolios in program evaluation: An investigation of interrater reliability. *American Journal of Evaluation*, 21(1), 65-80.
- Kablan, Z., Topan, B., & Erkan, B. (2013). The effectiveness level of material use in classroom instruction: A meta-analysis study. *Educational Sciences: Theory & Practice*. 13(3), 1638-1642.
- Kalaian, S., & Kasim, R. (2013). *A meta-analysis of the effectiveness of small-group instruction compared to lecture-based instruction in science, technology, engineering and mathematics (STEM) college classes*. Project Report. Retrieved from <https://arc.uchicago.edu/reese/projects/meta-analysis-effectiveness-small-group-instruction-compared-lecture-based-instructio>
- Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2006). *What does certification tell us about teacher effectiveness? Evidence from New York City*. (NBER Working Paper 12155). Cambridge, MA: National Bureau of Economic Research.

- Kane, T., Staiger, D., McCaffrey, D., Cantrell, S., Archer, J., Buhayar, S., Kerr, K., Kawakita, T. & Parker, D. (2012). Gathering Feedback for Teaching: Combining High-quality Observations with Student Surveys and Achievement Gains (Technical report), Measures of Effective Teaching Project. Seattle, WA: Bill & Melinda Gates Foundation.
- Kannapel, P. J., & Clements, S. K. (2005). *Inside the black box of high-performing high-poverty schools*. Lexington, KY: Prichard Committee for Academic Excellence.
- Kasper, G. (2000, March). *Four perspectives on L2 pragmatic development*. Plenary address given at the American Association of Applied Linguistics, Vancouver.
- Keller-Glaze, H., Horey, J., Nicely, K., Brusso, R., Nihill, M. M., & Cobb, M. G. (2013). *A practical decision guide for integrating digital applications and handheld devices into advanced individual training*. (Research Report 1967). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Kimball, S. M., White, B., Milanowski, A. T., & Borman, G. (2004). Examining the relationship between teacher evaluation and student assessment results in Washoe County. *Peabody Journal of Education*, 79(4), 54-78.
- Kirkpatrick, D. L. (1959a). Techniques for evaluating training programs. *Journal of American Society of Training Directors*, 13(11), 3-9.
- Kirkpatrick, D. L. (1959b). Techniques for evaluating training programs-Part 2: Learning. *Journal of the American Society of Training Directors*, 13(11), 21-26.
- Kirkpatrick, D. L. (1960a). Techniques for evaluating training programs-Part 3: Behavior. *Journal of the American Society of Training Directors*, 14(1), 13-18.
- Kirkpatrick, D. L. (1960b). Techniques for evaluating training programs-Part 4: Results. *Journal of the American Society of Training Directors*, 14(1), 28-32.
- Kirkpatrick, D. L. (1994). *Evaluating training programs: The four levels*. San Francisco, CA: Berrett-Koehler.
- Knowles, M. S. (1975). Self-directed learning: A guide for learners and teachers. New York: Association Press.
- Kolb, D. A. (1984). *Experiential Learning*,. Englewood Cliffs, NJ.: Prentice Hall.
- Kong, L. N., Qin, B., Zhou, Y. Q., Mou, S. Y., & Gao, H. M. (2014). The effectiveness of problem-based learning on development of nursing students' critical thinking skills: A systematic review and meta-analysis. *International Journal of Nursing Studies*, 51(3), 458-469.
- Kyriakides, L. (2005). Drawing from teacher effectiveness research and research into teacher interpersonal behavior to establish a teacher evaluation system: A study on the use of

student ratings to evaluate teacher behavior. *Journal of Classroom Interaction*, 40(2), 44-66.

Latham, G. P., Saari, L. M., Pursell, E. D., & Champion, M. A. (1980). The situational interview. *Journal of Applied Psychology*, 65(4), 422-427.

Lave, J. (1988). *Cognition in Practice: Mind, mathematics, and culture in everyday life*. Cambridge, UK: Cambridge University Press.

Lefrancois, G. (1999). *Psychology applied to teaching* (10th ed.). Belmont, CA: Wadsworth.

Levashina, J., Hartwell, C. J., Morgeson, F. P., & Campion, M. A. (2014). The structured employment interview: Narrative and quantitative review of the research literature. *Personnel Psychology*, 67, 241-293.

Lineberry, M., Bryan, E., Brush, T., Carolan, T., Holness, D., Salas, E., & King, H. (2013). Measurement and training of TeamSTEPPS dimensions using the Medical Team Performance Assessment Tool. *Joint Commission Journal on Quality and Patient Safety / Joint Commission Resources*, 39(2), 89-95.

Looi, C.K, Chen, W., & Ng, F.K. (2010). Collaborative activities enabled by GroupScribbles: An exploratory study of learning effectiveness. *Computers & Education*. 54(1), 14-26.

Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent Tutoring Systems and Learning Outcomes: A Meta-Analysis. *Journal of Educational Psychology*. Advance online publication. <http://dx.doi.org/10.1037/a0037123>

Marsh, H. W., & Dunkin, M. J. (1992). Students' evaluations of university teaching: A multidimensional perspective. In: J. C. Smart (Ed.) *Higher Education: Handbook of Theory and Research* (Vol. 8), (pp. 143-233), New York: Agathon Press.

Martinez, M. E. (2006). What is metacognition? *Phi Delta Kappan*, 87, 696-699.

McCaffrey, D. F., Lockwood, J. R., Koretz, D., Louis, T.A., & Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, 29, 67-101.

McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology*, 79(4), 599-616.

McEachin A. J., & Brewer, D. J. (2011). *Teacher intelligence: What is it and why do we care?* University of Southern California. Retrieved from <https://static1.squarespace.com/static/50c69b8de4b0c1ce70451157/t/50c6b636e4b0c1ce70456073/1355200054097/Intelligence+Draft+101511.pdf>

- Meirink, J. A., Meijer, P. C., Verloop, N., & Bergen, T. C.,M. (2009). How do teachers learn in the workplace? An examination of teacher learning activities. *European Journal of Teach Education*, 32(3) 209-224.
- Mertz, N. T. (2010). Teacher selection and school leader effects. *Journal of School Leadership*, 20(2), 184-207.
- Metzger, S. A., & Wu, M. (2008). Commercial teacher selection instruments: The validity of selecting teachers through beliefs, attitudes, and values. *Review of Educational Research*, 78(4), 921-940.
- Meyers, R. (2010). *Army reserve instructors' perceptions regarding the effectiveness of experiential learning model in teaching mid-level Army reserve officers*. ProQuest LLC: Dissertation, University of South Dakota.
- Meyers, S., & Lester, D. (2013). *The effects of situated learning through a community partnership in a teacher preparation program*. Retrieved from <http://sgo.sagepub.com/content/3/3/2158244013497025>
- Molenaar, I., Chie, M.M., Sleegers, P., & van Boxtel, C. (2011). Scaffolding of small groups' metacognitive activities with an avatar. *Computer-Supported Collaborative Learning*. 6, 601-624.
- Morrison, J. E. & Fletcher, J. D. (2002). *Cognitive Readiness* (IDA Paper P-3735). Alexandria, VA: Institute for Defense Analysis.
- Morsh, J. E., & Wilder, E. W. (1954). *Identifying the effective instructor: A review of the quantitative studies (Report No. AFPTRC-TR-54-44)*. Chanute AFB, IL: Air Force Personnel and Training Research Center.
- Murad, M. H., Coto-Yglesias, F., Varkey, P., Prokop, L. J., & Murad, A. L. (2010). The effectiveness of self-directed learning in health professions education: A systematic overview. *Medical Education*, 44, 1057–1068. doi: 10.1111/j.1365-2923.2010.03750.x
- Naismith, L., Lonsdale, P., Vavoula, G., & Sharples, M. (2004). *Report 11: Literature review in mobile technologies and learning*. Retrieved from: <http://archive.futurelab.org.uk/resources/publications-reports-articles/literature-reviews/Literature-Review203>
- National Education Association. (2010). *Teacher assessment and evaluation: The National Teacher Association's framework for transforming education systems to support effective teaching and improve student learning*. Retrieved from http://www.nea.org/assets/docs/HE/TeachrAssmntWhtPaperTransform10_2.pdf
- Naval Education and Training Command. (2010). *Naval education and training center instruction 1500.5, Instructor preparation, qualification, certification, and evaluation program* (NTEC Instruction 1500.5). Pensacola, FL: Author.

- Ng, T. W. H., & Feldman, D. C. (2009). How broadly does education contribute to job performance? *Personnel Psychology*, 62, 89-134.
- Nielsen, T. (2008). Implementation of learning styles at the teacher level. *Education & Training*, 50, 155–166. doi: 10.1108/00400910810862137
- Niemiec, R.P. & Walberg, H.J. (1992). The effect of computers on learning. *International Journal of Education*, 17, 99–108.
- Oliva, M., Mathers, C., & Laine, S. (2009, March). Effective evaluation. *Principal Leadership*, 10, 16-21.
- Painter, B. (2001). Using teaching portfolios. *Educational Leadership*, 2, 31-34.
- Paulsen, M.B. (2002). Evaluating teaching performance. *New Directions for Institutional Research*, 114, 5-18.
- Pearce, J., Mann, M.K., Jones, C., van Buschbach, S., Olff, M., & Bisson, J. I., (2012). The most effective way of delivering a train-the-trainers program: A systematic review. *Journal of Continuing Education in the Health Professions*, 32(3), 215-226.
- Phillips, J. J. (1983). *Handbook of training and evaluation methods*. (1st ed.). Houston, TX: Gulf Publishing.
- Pleban, R. J., Blankenbeckler, P. N., Wampler, R. Ll, Dlubac, M. D., & Perdomo, B. (2013). *Comparison of direct instruction and problem centered instruction for Army institutional training*. (Research Report 1966). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Pressley, M., Wharton-McDonald, R., Allington, R., Block, C. C., Morrow, L., Tracey, D., et al. (2001). A study of effective grade-1 literacy instruction. *Scientific Studies of Reading*, 5(1), 35–58.
- Puntambekar, S., & Kolodner, J. L. (2005). Distributed scaffolding: Helping students learn science by design. *Journal of Research in Science Teaching*, 42(2), 185-217.
- Reed, D. Liu, A. Y., Kleinman, R., Mastri, A., Reed, D., Sattar, S., & Ziegler, J. (2012). An effectiveness assessment and cost-benefit analysis of registered apprenticeship in 10 states. *Final Report submitted to Department of Labor Employment and Training Administration*. Oakland, CA: Mathematica Policy Research.
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417-458. Retrieved from <http://www.econ.ucsb.edu/~jon/Econ230C/HanushekRivkin.pdf>
- Rogelberg, S. (2007). *Encyclopedia of industrial and organizational psychology*. Thousand Oaks, CA: SAGE Publications, Inc.

- Roh, K. H., & Park, H-A. (2010). A meta-analysis on the effectiveness of computer-based education in nursing. *Health Information Research*, 16(3), 149-157.
- Root, L. S. (1987). Faculty evaluation: Reliability of peer assessments of research, teaching, and service. *Research in Higher Education*, 26(1), 71-84.
- Roth, P. L., Bobko, P., & McFarland, L. A. (2005). A meta-analysis of work sample test validity: Updating and integrating some classic literature. *Personnel Psychology*, 58(4), 1009-1037.
- Schaefer, P. S., & Dyer, J. L. (2012). Bridging the gap between adaptive training research and Army practice. *Military Psychology*, 24, 194-219.
- Schalock, D. (1979). Research on teacher selection. *Review of Research in Education*, 7, 364-417.
- Schatz, S., Bartlett, K., Burley, N., Dixon, D., Knarr, K., & Gannon, K. (2012). *Making good instructors great: USMC cognitive readiness and instructor professionalism*. Paper presented at the Interservice/Industry Training, Simulation and Education Conference, Orlando, FL.
- Schmitt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124(2), 262-274.
- Seemiller, C. (2014). *The student leadership competencies guidebook*. San Francisco, CA: Jossey-Bass.
- Seldin, P. (2006). *Evaluating faculty performance*. San Francisco, CA: Jossey-Bass.
- Shandler, D. (2000). *Competency and the Learning Organization*. Mississauga, Ontario: Crisp Learning.
- Sharan, S., & Shaulov, A. (1990). Cooperative learning, motivation to learn, and academic achievement. In Sharan, Shlomo, (Ed.), *Cooperative learning: Theory and research* (pp.173-202) New York: Praeger.
- Sharma, P. (2010). Blended learning. *English Language Teachers Journal*, 64(4), 456-458.
- Sheftal, M. S. (2000). *Teacher expectations, teacher efficacy, and student achievement*. Athens, GA: University of Georgia.
- Sitzmann, T., & Ely, K. (2011). A meta-analysis of self-regulated learning in work-related training and educational attainment: What we know and where we need to go. *Psychological Bulletin*, 137 (3), 421 – 442.
- Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology*, 64, 489-528.

- Slavin, R. E. (1995). *Cooperative learning: Theory, research, and practice* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Smith, T. M., & Ingersoll, R. (2004). What are the effects of induction and mentoring on beginning teacher turnover? *American Educational Research Journal*, 41(3), 681-714.
- Smith-Jentsch, K. A., Cannon-Bowers, J. A., Tannenbaum, S. I., & Salas, E. (2008) Guided team self correction: Impacts on team mental models, processes and effectiveness. *Small Group Research*, 39(3), 303-327.
- Starenko, M., Vignare, K., & Humbert, J. (2007). Chapter 8: Enhancing Student Interaction and Sustaining Faculty Instructional Innovations through Blended Learning. In A. Picciano & C. Dzieban (Eds), *Blended Learning: Research perspectives* (pp.150-179) Rochester, NY: Rochester Institute of Technology.
- Stronge, J. H. (1997). Improving schools through teacher evaluation. In J. H. Stronge (Ed.), *Evaluating teaching: A guide to current thinking and best practice* (pp. 1-23), Thousand Oaks, CA: Corwin Press.
- Stronge, J.H, & Hindman, J.L. (2006). *The Teacher Quality Index: a Protocol for Teacher Selection*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Teaching.org (n.d.) *Become a teacher*. Retrieved from <https://www.teach.org/>
- The Chronicles of Higher Education, n.d. *Find Jobs*. Retrieved from <https://chroniclevitae.com/>
- Thorton, G. C., & Gibbons, A. M. (2009). Validity of assessment centers for personnel selection. *Human Resource Management Review*, 19, 169-187.
- Tolliver, R. (2014). Boldly transforming leadership development. Retrieved from http://www.army.mil/article/119271/Boldly_Transforming_Leadership_Development/
- Tucker, P.D., Stronge, J.H., Gareis, C.R., & Beers, C.S. (2003). The efficacy of portfolios for teacher evaluation and professional development: Do they make a difference? *Educational Administration Quarterly*, 39(5), 572-602.
- Turpen, C., Henderson, C. & Dancy, M. (2012). Faculty Perspectives about Instructor and Institutional Assessments of Teaching Effectiveness. *AIP Conference Proceedings*, 1413(1), 371-371.
- U.S. Army Training and Doctrine Command (2013a). *Noncommissioned officer education system instructor development and recognition program*. (TRADOC Regulation 600-21). Fort Eustis, VA: Author.
- U.S. Army Training and Doctrine Command. (2013b). *Staff and faculty development* (TRADOC Pamphlet 350-70-3). Fort Eustis, VA: Author.

- U.S. Army Training and Doctrine Command. (2013c). *Army learning policies and systems* (TRADOC Regulation 350-70). Fort Eustis, VA: Author.
- U.S. Army Training and Doctrine Command. (2011). *The U.S. Army learning concept for 2015* (TRADOC Pamphlet 525-8-2). Fort Monroe, VA: Author.
- U.S. Coast Guard Force Readiness Command. (2011). *Standing operating procedures for the Coast Guard's training system, Volume 3: Evaluation*. Norfolk, VA: Author.
- U.S. Department of the Air Force (2012). *Faculty development and master instructor programs* (AETC Instruction 36-2202). Randolph Air Force Base, TX: Author.
- U.S. Department of the Army (2005). *Army Regulation 614-200, Enlisted assignments and utilization management* (U.S. Army Regulation 614-200). Washington, D.C.: Author
- U.S. Department of the Army (2010). *The Army School System*. (Army Regulation 350-18). Fort Monroe, Virginia: Author.
- U.S. Department of the Army (2011a). *The U.S. Army learning model for 2015*. (TRADOC Pamphlet 525-8-2). Fort Monroe, Virginia: Author.
- U.S. Department of the Army (2011b). *The U.S. Army training concept: 2012-2020* (TRADOC Pamphlet 525-8-3). Fort Monroe, Virginia: Author.
- U.S. Department of the Army. (2014). *Personnel evaluation: Evaluation reporting system* (U.S. Army Pamphlet 623-3). Washington, D.C.: Author.
- U.S. Department of the Navy (2011). *Management of Marine Corps formal schools and training detachments* (U.S. Marine Corps Order 1553.2A and 1553.2B). Washington, D.C.: Author
- Utley, B. L. (2006). Effects of situation learning on knowledge gain of instructional strategies of students in a graduate level course. *Teacher Education and Special Education*, 29(1), 69-82.
- VanSickle, R. L. (1986). A quantitative review of research on instructional simulation gaming: A twenty-year perspective. *Theory and Research in Social Education*, 14(3), 245-264.
- Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research*. 34(3), 229-243.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Wachtel, H. K. (1998). Student evaluation of college teaching effectiveness: A brief review. *Assessment & Evaluation in Higher Education*, 23(2), 191-212.

- Wade, R. (1985). What Makes a Difference in Inservice Teacher Education? A Meta-Analysis of Research. *Educational Leadership*, 42(4), 48-54.
- Walker, A., & Leary, H. (2009). A problem based learning meta analysis: Differences across problem types, implementation types, disciplines, and assessment levels. *Interdisciplinary Journal of Problem-Based Learning*, 3(1), 12-43.
- Waxman, H. C., Lin, M., & Michko, G. M. (2003). *A meta-analysis of the effectiveness of teaching and learning with technology on student outcomes*. Naperville, IL: Learning Point Associates. Retrieved from <http://www.ncrel.org/tech/effects2/>
- Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research*, 73(1), 89-122.
- Webb, N., & Palincsar, A. (1996). Group processes in the classroom. In D.C. Berliner & R.C. Calfee (Eds.), *Handbook of educational psychology* (pp. 841-873). New York: Simon & Schuster.
- Webster, W. J. (1988). Selecting effective teachers. *Journal of Educational Research*, 81(4), 245-253.
- Wenglinsky, H. (2002, February 13). How schools matter: The link between teacher classroom practices and student academic performance. *Education Policy Analysis Archives*, 10(12). Retrieved from <http://epaa.asu.edu/epaa/v10n12/>
- White, C. R., Carson, J. L., & Wilbourn, J. M. (1991). Handgun marksmanship training: Evaluation of an advanced marksmanship trainer. *Performance Improvement Quarterly*, 4(3), 63-73.
- Wiliam, D. (2006). Formative assessment: Getting the focus right. *Educational Assessment*, 11, 283-289.
- Wilkerson, D. J., Manatt, R. P., Rogers, M. A., & Maughan, R. (2000). Validation of student, principal and self-ratings in 360° feedback for teacher evaluation. *Journal of Personnel Evaluation in Education*, 14(2), 179-192.
- Wilson, S. M., & Berne, J. (1999). Chapter 6: Teacher Learning and the Acquisition of Professional Knowledge: An Examination of Research on Contemporary Professional Development. *Review of Research in Education*, 24, 173-209.
- Winter, P. A. (1995). Facts and fiction about teacher selection: Insights from current research findings. *The High School Journal*, 79(1), 21-24.
- Wise, A. E., Darling-Hammond, L., & Berry, B. (1987). *Effective teacher selection: From recruitment to retention (R-3462-HIE/CSTP)*. Santa Monica, CA: RAND Corporation.

Worrell, F. C., & Kuterbach, L. D. (2001). The use of student ratings of teacher behaviors with academically talented high school students. *Journal of Secondary Gifted Education*, 14(4), 236-247.

Yoon, S. W., & Lim, D. H. (2007). Strategic blending: A conceptual framework to improve learning and performance. *International Journal on E-Learning*, 6, 475 – 489.

Yoon, S. W., Duncan, T., Lee, S. W-Y., Scarloss, & Shapley. (2007). Reviewing the evidence on how professional development affects student achievement. (REL 2007- No. 33), *National Center for Education Evaluation and Regional Assistance*. Retrieved from ies.ed.gov/ncee/edlabs/regions/southwest/pdf/REL_2007033.pdf